NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

THESIS

CURRENT ISSUES CONCERNING
RELIABILITY ESTIMATION IN
OPERATIONAL TEST AND EVALUATION

by

Timothy P. Anderson

September 1994

Thesis Advisors: Donald P. Gaver
Patricia A. Jacobs

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704

1 AGENCY USE ONLY *(Leave blank)*

2 REPORT DATE September 1994

3. REPORT TYPE AND DATES COVERED Master's Thesis

4 TITLE AND SUBTITLE CURRENT ISSUES CONCERNING RELIABILITY ESTIMATION IN OPERATIONAL TEST AND EVALUATION (U)

5. FUNDING NUMBERS

6 AUTHOR(S) Anderson, Timothy P.

7 PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey CA 93943-5000

8 PERFORMING ORGANIZATION REPORT NUMBER

9 SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11 SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE
*A

13. ABSTRACT *(maximum 200 words)*

There are several perennial issues concerning reliability estimation in Operational Testing and Evaluation (OT&E). Two of these include, "how should one model the underlying failure distribution of a continuous-time system," and, "how can a testing agency use information from DT in order to reduce OT resource requirements."

In the former issue, some OT&E analysts have questioned whether or not the exponential failure distribution should be used in all cases for continuous-time systems, and have suggested the Weibull distribution as an alternative in some instances. In the latter, the notion of combining DT and OT data has been an anathema to those involved in OT&E, however, with ever-tightening military budgets, it may be time to revisit the issue.

First, this thesis compares the exponential and Weibull failure distributions in terms of the amount of test time needed to demonstrate, to a given level of confidence, that the true MTTF of a system is at least as large as the minimum acceptable value, and also in terms of the actual confidence level associated with the lower confidence level procedure when the system has an increasing (or decreasing) failure rate function. Second, the thesis examines the behavior of an estimator for the relationship between DT and OT failure data using a Monte Carlo simulation. Finally, the thesis introduces a hierarchical Bayes approach for the estimation of the relationship between DT and OT failure data when a gamma prior distribution is assumed.

14. SUBJECT TERMS Reliability estimation, Weibull distribution, Exponential distribution, Developmental testing, Operational test and evaluation.

15. NUMBER OF PAGES 65

16. PRICE CODE

17 SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500 Standard Form 298 (Rev 2-89)

Approved for public release; distribution is unlimited.

CURRENT ISSUES CONCERNING RELIABILITY ESTIMATION IN OPERATIONAL TEST AND EVALUATION

by

Timothy P. Anderson
Lieutenant, United States Navy
B.S. Industrial and Operations Engineering, University of Michigan, 1987

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1994

Author: Timothy P. Anderson
Timothy P. Anderson

Approved by: Donald P. Gaver
Donald P. Gaver, Thesis Advisor

Patricia A. Jacobs
Patricia A. Jacobs, Thesis Advisor

W. Max Woods
W. Max Woods, Second Reader

Peter Purdue, Chairman
Department of Operations Research

ABSTRACT

There are several perennial issues concerning reliability estimation in Operational Testing and Evaluation (OT&E). Two of these include, "how should one model the underlying failure distribution of a continuous-time system," and, "how can a testing agency use information from DT in order to reduce OT resource requirements."

In the former issue, some OT&E analysts have questioned whether or not the exponential failure distribution should be used in all cases for continuous-time systems, and have suggested the Weibull distribution as an alternative in some instances. In the latter, the notion of combining DT and OT data has been an anathema to those involved in OT&E, however, with ever-tightening military budgets, it may be time to revisit the issue.

First, this thesis compares the exponential and Weibull failure distributions in terms of the amount of test time needed to demonstrate, to a given level of confidence, that the true MTTF of a system is at least as large as the minimum acceptable value, and also in terms of the actual confidence level associated with the lower confidence level procedure when the system has an increasing (or decreasing) failure rate function. Second, the thesis examines the behavior of an estimator for the relationship between DT and OT failure data using a Monte Carlo simulation. Finally, the thesis introduces a hierarchical Bayes approach for the estimation of the relationship between DT and OT failure data when a gamma prior distribution is assumed.

THESIS DISCLAIMER

The reader is cautioned that computer programs developed in this research may not have been exercised for all cases of interest. While effort has been made, within the time available, to ensure that the programs are free of computational and logic errors, they cannot be considered validated. Any application of these programs without additional verification is at the risk of the user.

TABLE OF CONTENTS

I. INTRODUCTION 1

II. COMPARISON OF EXPONENTIAL AND WEIBULL DISTRIBUTIONS 5

A. EXPONENTIAL MODEL 5

B. ON CONFIDENCE LEVELS 8

C. WEIBULL MODEL 9

D. TEST TIME NEEDED TO DEMONSTRATE A THRESHOLD 12

E. USE OF EXPONENTIAL MODEL ON A WEIBULL PROBLEM 14

F. CONCLUSIONS 16

III. ON COMBINING DT AND OT FAILURE DATA 18

A. BACKGROUND AND PROBLEM DESCRIPTION 18

B. THE GAVER-JACOBS MODEL 19

C. ANALYSIS OF THE GAVER-JACOBS MODEL 21

D. RESULTS OF THE GAVER-JACOBS MODEL ANALYSIS 39

E. A HIERARCHICAL BAYES MODEL FOR ESTIMATION OF K 40

IV. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS 46

A. SUMMARY 46

B. CONCLUSIONS 47

C. RECOMMENDATIONS FOR FURTHER STUDY 48

APPENDIX A - ALTERNATIVE METHOD OF DERIVING TIME TO TEST 51

APPENDIX B - CODE FOR RANDOM POISSON ARRIVAL GENERATOR 53

LIST OF REFERENCES 55

INITIAL DISTRIBUTION LIST 56

EXECUTIVE SUMMARY

There are several perennial issues concerning reliability estimation in Operational Testing and Evaluation (OT&E). One of these is, "how should one model the underlying failure distribution of a continuous-time system," and another is, "how can a testing agency use information from Developmental Testing (DT) in order to reduce Operational Testing (OT) resource requirements."

The purpose of this thesis is to (1) investigate the behavior of a statistical procedure used to determine the length of time needed to test a continuously operating system when the system under test is assumed to have an exponential time to failure distribution, but when in fact the system does not have an exponential time to failure distribution and rather may have a time to failure distribution that more closely resembles a Weibull distribution; (2) to examine the behavior of a maximum likelihood estimator for the relationship between DT and OT failure rates using a Monte Carlo simulation; and (3) to introduce a hierarchical Bayes method for estimation of the relationship between the DT and OT failure rates. In the first part of the thesis, the parameters of the procedure are chosen based on the assumption of an exponential time to failure distribution. The behavior of the procedure is then investigated under the assumption that the time between failures are independent, having a Weibull distribution with the same mean time to failure. The intention is to test whether or not use of the Weibull distribution will result in fewer test hours to verify that a system under test has a mean time to failure which meets or exceeds a minimum acceptable value at a given level of confidence, and also to test whether or not the OTAs are misstating the confidence they report in their operational testing results when the exponential distribution is used to model a system whose true failure rate function is that of a Weibull distribution. In the second part of the thesis, a maximum likelihood estimator is examined in terms of the consistency and variability of the OT failure rate derived using that estimator compared to other methods of establishing the OT failure rate. Finally, in the third part of the thesis, a hierarchical Bayes model is proposed for the ultimate purpose of being able to describe the mean and variance associated with a random relationship between DT and OT failure rates.

The comparison of the exponential and Weibull failure models suggests that operational testing agencies can, under certain circumstances, reduce the amount of time needed to demonstrate that the *mean time to failure* (MTTF) of a continuously operating system is at least as large as the *minimum acceptable value* (MAV) if the system's true underlying failure distribution is Weibull with shape parameter $\beta \neq 1$. However, the shape parameter is seldom, if ever, known or even estimable without expending valuable test resources, so it is unlikely that an *operational testing agency* (OTA) could actually save any time or resources with this methodology. Moreover, from the perspective of confidence levels, the analysis suggests that confidence in one's result is not sacrificed when the exponential distribution is used to model a system which has a non-constant failure rate, since it appears that use of the exponential distribution on a truly Weibull system seems to cause the OTA to overestimate the number of hours needed to test the MAV for the MTTF to a given confidence level, and also to underestimate the confidence it reports in its results under the incorrect assumption of an exponential distribution. The thesis also points out, however, that these results may not be true in general. Therefore, it is concluded that there is currently little to be gained by attempting to model the time between system failures with the Weibull distribution as previously suggested, particularly if very few observations of time to failure are available.

The second analysis indicates that the maximum likelihood method for estimating a system's OT failure rate given that system's DT failure rate as well as failure data on previous similar systems produces a reasonable estimate of the actual OT failure rate (when the OT failure rate behaves as expected). In fact, the estimate provided by the maximum likelihood method appears to be nearly as good as direct observation of the OT failure rate obtained through actual testing. Thus, if it can be shown that inferences about the failure rate of a system can in fact be made on the basis of previous similar systems, then the maximum likelihood method of predicting a new system's OT failure rate from previous DT failure data has some potential utility. In particular, this OT failure rate prediction could be used in ways which would aid decisions as to when to begin OT as well as to augment follow-on OT in an effort to reduce needed testing resources.

The third analysis is incomplete. It was hoped that the hierarchical Bayes method given would provide a means by which to estimate the relationship between the DT and OT failure rates when the relationship is assumed to be a random variable with some prior distribution. Regretfully, the equations derived for the maximum likelihood estimators for the parameters of the gamma prior have proved to be difficult to solve satisfactorily. If a solution to these equations were available, it could be used to predict future failure rates by means of a Bayesian calculation. Investigation of such a procedure is left for a later time.

I. INTRODUCTION

There are several perennial issues concerning reliability estimation in Operational Test and Evaluation (OT&E). In continuous systems, such as radios, radars and other items that operate continuously while in use, one perennial issue is "how should one model the underlying failure distribution of the system?" Another is "what can be done to reduce the amount of testing resources needed during OT&E?"

Historically, OT&E organizations, such as Commander, Operational Test and Evaluation Force (COTF), have used the exponential distribution to model the behavior of the reliability of a continuous-time type system. The assumption of an exponential time to failure distribution results in a required test period of three times a given *minimum acceptable value* (MAV) for the 80 percent lower confidence limit for the *mean time to failure* (MTTF). As long as fewer than two failures occur during this time, then COTF will report 80 percent confidence that the true MTTF of the system is at least as large as the MAV (COTF, para. 204.a.(3), 1992). COTF analysts have questioned whether or not the exponential distribution is the best underlying failure model to use, and have considered investigating alternative distributions (Madson, 1993). One possible alternative in question is the Weibull distribution, which contains the exponential distribution as a special case. The main difference between the two underlying failure distributions is that the exponential distribution has a constant failure rate function throughout the lifetime of the system, and the Weibull distribution has a changing failure rate function (either increasing, constant or decreasing) during the lifetime of the system. The idea is that the Weibull distribution might provide a "better fitting" model for certain systems which might then reduce the amount of test time needed to demonstrate, to a given confidence level, that the true MTTF of a system is at least as large as the MAV.

A third perennial issue is that of combining Developmental Testing (DT) failure data and Operational Testing (OT) failure data in a way that would reduce the overall amount of resources needed to establish the reliability of a system. For example, Gaver and Jacobs (1994, p. 1) describe the following problem setting:

> A test and evaluation analyst has in his possession historical (time-between-failure or equivalent) data for both the...DT...and...OT...phases of a number of projects that are roughly comparable: they are all C3/I systems, perhaps. It is believed that there may be a usefully exploitable relationship between the DT and OT data. If this relationship could be quantified then perhaps it could be used to augment DT data for a new project of the same type, thereby obtaining some anticipation of the OT data for that new project...

Gaver and Jacobs (1994, p. 5) go on to devise an estimator for the relationship between the DT data and OT data and describe its application thusly:

> The ultimate application of the above (relationship) is to (a) anticipate failure patterns during Operational Testing, given data on failures during Developmental Testing; this might aid in the decision as to when to begin OT; and (b) to strengthen, and reduce uncertainty of, the post-DT estimates of...(system reliability)...by incorporating the DT data.

Of course, in the preceding example, it is assumed that the relationship between the DT and OT failure rates is constant across all "similar" systems. It is far more likely that the relationship follows some probability distribution. Thus, a hierarchical approach for the description of this relationship may be appropriate. One such approach is outlined in the last part of this thesis.

The notion of combining DT and OT data has historically been an anathema to those involved in OT&E. The primary reason for this is that the Operational Testing Agencies (OTAs) have, rightly so, desired to be completely independent of the Developmental Agencies (DAs). However, with the tight budgets facing all military organizations of late, it may be time to re-visit the idea of using previous DT data to supplement OT data in an effort to reduce the amount of testing resources needed to establish a system's reliability in OT&E.

The first goal of this thesis is to compare, through a sensitivity analysis, the exponential and Weibull failure distributions in terms of the amount of test time needed to demonstrate, to a given level of confidence, that the true MTTF of a system is at least as large as the MAV, and to find the actual confidence level associated with the lower confidence limit procedure when the system has an increasing (or decreasing) failure rate function. The former tests whether or not use of the Weibull distribution will result in fewer test hours to establish reliability with a given confidence level. The latter tests whether or not the OTA is overstating (or understating) the confidence it reports in its results when the exponential distribution is assumed but the system's true failure rate follows a Weibull distribution.

The second goal of this thesis is to examine the behavior of the estimator for the relationship between DT and OT data devised by Gaver and Jacobs (1994) using a Monte Carlo simulation. The behavior to be examined includes the consistency and variability of the OT failure rate derived using this estimator compared to those of other methods of establishing the OT failure rate, including direct observation of the OT data and a weighted average between direct observation and the Gaver-Jacobs estimate of the OT failure rate.

The third goal of this thesis is to introduce a hierarchical Bayesian approach for the estimation of the relationship between DT and OT data using a gamma prior distribution.

Note that the methods used for borrowing DT information to assist in predicting OT information, or actual later field experience, can also be used in other situations. For example, suppose an upgrade of an existing system is to be evaluated. Then actual data from the existing system may be scaled to the new system in a manner analogous to that already described for DT-to-OT scaling. Note also that the general procedure described can be adapted to other operational parameters. An example might be an effectiveness parameter such as projectile dispersion around an aim point, technically speaking, a variance parameter. Under current conditions of declining resources such an approach could be of considerable value in enhancing the accuracy of many types of test data.

II. COMPARISON OF EXPONENTIAL AND WEIBULL DISTRIBUTIONS

A. EXPONENTIAL MODEL

One of the simplest and most popular ways to model the distribution of the random time to failure of a continuous time system is with the exponential distribution. Among the reasons for the popularity of the exponential distribution are its simplicity, ease of use, and the memoryless property which says that the probability of failure within the next t time units is independent of the frequency of any previous failures.

Let T = random time to failure of a system,
t = length of time system is in operation,
γ = failure rate.

The time to failure, T, of a system follows an exponential failure distribution if the probability density function (PDF) is

$$f(t) = \gamma e^{-\gamma t}, t \geq 0 \tag{2.1}$$

and if its cumulative distribution function (CDF) is

$$\begin{aligned} F(t) &= 1 - e^{-\gamma t}, t \geq 0 \\ &= 0, t < 0 \end{aligned} \tag{2.2}$$

If the times between system failures are independent and have the exponential distribution specified above, then the system's mean time to failure (MTTF) is

$$\text{MTTF} = E[T] = \frac{1}{\gamma} \tag{2.3}$$

and its failure rate function is

$$h(t) = \frac{f(t)}{1 - F(t)} = \gamma. \tag{2.4}$$

This distribution is used by the analysts and Operational Test Directors (OTDs) at COTF primarily to determine the length of time needed to test a given continuous time system in order to demonstrate that the true MTTF of the system is greater than or equal to the MAV for the MTTF with approximately known confidence. The basic assumption associated with the exponential distribution is that the system under scrutiny exhibits steady state performance; that is, the time between failures are independent and have a common exponential distribution. In other words, it is assumed that the failure rate function, h(t), is constant over time: there is no "reliability growth," or degradation.

COTF's usual policy is to conduct sequential tests. That is, they will test one system for a pre-determined length of time. If it fails before the time on test expires then they will repair the system and continue the test. If the repaired system fails before the remaining time on test then the system fails the test, else it passes. In other words, if there are zero or one failures during the time on test, then the system passes (COTF, para. 602.a, 1992).

Let α = confidence level,
θ = MAV for the MTTF,
T_2 = time to 2nd failure,
$N(t)$ = number of failures to occur in $[0, t]$.

The time, t, needed for $100(\alpha)$ percent confidence, with no more than one failure allowed during testing, required to demonstrate that the system meets the MAV for the MTTF, θ, is determined as follows:

$$1-\alpha \geq P(T_2 > t) = P(N(t) \leq 1)$$

$$1-\alpha \geq \sum_{i=0}^{1} \frac{\left(\frac{t}{\theta}\right)^i e^{-\left(\frac{t}{\theta}\right)}}{i!} = e^{-\left(\frac{t}{\theta}\right)} + \left(\frac{t}{\theta}\right) e^{-\left(\frac{t}{\theta}\right)}. \tag{2.5}$$

Standard procedure at COTF is to set the confidence level at 80 percent and to allow for a maximum of one failure to occur during the operational test. Solution of equation (2.5) with $\alpha = 0.80$ gives a time to test, $t = 3\theta$.[1] In other words, COTF will plan to test a continuous time system for a period of time which is three times greater than the MAV for the MTTF, and if the system has fewer than two failures in this time, then the system passes the test at the 80 percent confidence level. Note that other combinations of test times and acceptable numbers of failures can yield the same confidence, but the above has been chosen as a practical compromise.

For example, if the MAV for the MTTF is 1000 hours, COTF will test the system for 3000 hours and allow at most one failure to occur. If the system passes this test, COTF will report that it is 80 percent confident that the system's true MTTF meets or exceeds the MAV.

Implicit in the preceding model, however, is the notion that if the expected value of the time between failures for one system is greater than or equal to the expected value of the time between failures for a second system, then the probability of having J or fewer failures in a renewal process observed for a fixed time for the first system is at least as large as that for the second system given the same fixed time.

Let X = time to 1st failure of system 1, $X \sim \exp(\gamma_1)$,
Y = time to 1st failure of system 2, $Y \sim \exp(\gamma_2)$, $\gamma_1 \le \gamma_2$.

Proposition 1: If $E[X] \ge E[Y]$ then $P(N_X(t) \le 1) \ge P(N_Y(t) \le 1)\ \forall\ t$.

Proof: $\gamma_1 \le \gamma_2 \quad \Rightarrow \quad E[X] = 1/\gamma_1 \ge E[Y] = 1/\gamma_2$

$\gamma_1 \le \gamma_2 \quad \Rightarrow \quad \overline{F}_X(t) = e^{-\gamma_1 t} \ \ge\ \overline{F}_Y(t) = e^{-\gamma_2 t} \quad \forall\ t \ge 0$

---

[1] An interesting, and simpler, alternative method for determining t exists, involving the chi-square distribution. See Appendix A for details of the alternative method.

Let $N_X(t)$ = the number of renewals that occur in a renewal process having inter-arrival distribution $F_X(t)$,
$N_Y(t)$ = the number of renewals that occur in a renewal process having inter-arrival distribution $F_Y(t)$.

then $\bar{F}_X(t) \geq \bar{F}_Y(t) \Rightarrow N_X(t) \leq_{ST} N_Y(t) \ \forall\ t \geq 0$ (Ross, p.257, 1983),

which means that $P(N_X(t) \leq 1) \geq P(N_Y(t) \leq 1)$. *Q.E.D.*

Note that Proposition 1 may not be true for all eligible distributions. However, the fact that it is true for the exponential distribution enables one to use the COTF procedure.

B. ON CONFIDENCE LEVELS

One might ask at this point how COTF defines "level of confidence." As used in equation (2.5), the confidence level is actually the "producer's risk." When COTF desires to test a system's MTTF with 80 percent confidence, this means they are willing to assume a 20 percent (or smaller) risk of incorrectly recommending a system for production based on the results of their test.

Let $N_X(t)$ and $N_Y(t)$ denote, respectively, the number of renewals to occur in two renewal processes having distribution $F_X(t)$ and $F_Y(t)$. Ross (p. 257, 1983) shows that if $\bar{F}_X(t) \geq \bar{F}_Y(t)$ then $P(N_X(t) \leq 1) \geq P(N_Y(t) \leq 1)$. Since $\bar{F}_X(t) \geq \bar{F}_Y(t)$ implies that $E[X] \geq E[Y]$, the result shows that under this additional assumption, if the true MTTF of the system is less than the MAV, θ, and the desired level of confidence is α, then Ross's result implies that $P(N(t) \leq 1) < 1 - \alpha$ and $P(N(t) > 1) > \alpha$.

In other words, if COTF conducted the same test many times on a system with exponential times between failures, and if the true MTTF were actually less than the MAV, θ, then the observed time for the second failure to occur, T_2, would fall below 3θ more than 80 percent of the time (in which case COTF would correctly reject the system). Consequently, less than 20 percent of the time the observed time for the second failure to occur, T_2, would lie above 3θ (in which case they would incorrectly accept the system). Symbolically, COTF desires:

$$P(\text{accept the system} \mid \text{true MTTF} < \text{MAV}) \leq .20.$$

There is an interesting implication to this requirement. In order to have any appreciable chance of passing a test set up according to COTF standards, the DA must build its system such that its true MTTF is much greater than the MAV (Keller, 1993). If the system's MTTF is equal to the MAV, then it has only a 20 percent chance of passing the test.

Note carefully: the above does <u>not</u> say that if the test leads to acceptance, i.e., that fewer than two failures are observed in three times the MAV, then the probability is 80 percent that the true MTTF is greater than the MAV. This statement is specifically not allowed in the present framework.

C. WEIBULL MODEL

An alternative, and more general, way to model the probability of failure of a continuous system is with the Weibull distribution. The Weibull distribution is useful when the system under scrutiny displays either an increasing or decreasing failure rate over time. An example of a system with an increasing failure rate function would be one that "wears out" over time. A system that exhibits "wear out" would tend to have a higher failure rate in old age than an identical system which is younger. An example of a system with a decreasing failure rate function would be one that "wears in" over time. Certain electronic devices may exhibit "wear in" tendencies, at least initially.

Let T = random time to failure of a system,
t = length of time system is in operation,
λ = scale parameter (characteristic life),
β = shape parameter.

The time to failure, T, of a system follows a Weibull failure distribution if the probability density function (PDF) is

$$f(t) = \beta\lambda^{\beta}t^{\beta-1}e^{-(\lambda t)^{\beta}}, \; t \geq 0, \beta > 0, \lambda > 0 \tag{2.6}$$

and if its cumulative distribution function (CDF) is

$$\begin{aligned} F(t) &= 1 - e^{-(\lambda t)^{\beta}}, \; t \geq 0, \beta > 0, \lambda > 0 \\ &= 0, \; t < 0. \end{aligned} \tag{2.7}$$

If the times between system failures are independent and have the Weibull distribution specified above, then the mean time to failure (MTTF) of the system is

$$\text{MTTF} = E[T] = \frac{\Gamma\left(\frac{1}{\beta}\right)}{\beta\lambda} \tag{2.8}$$

and its failure rate function is

$$h(t) = \frac{f(t)}{1-F(t)} = \beta\lambda^{\beta}t^{\beta-1}. \tag{2.9}$$

This distribution has been considered as a replacement for the exponential distribution with the hope that it might more precisely model many of the systems under test by COTF. In this model, λ is analogous to the exponential failure rate, γ. The shape parameter, β, determines whether the system's failure rate is increasing, decreasing, or constant over time. When β is less than one, the failure rate function is decreasing, when β is greater than one, the failure rate function is increasing, and when β is equal to one, the failure rate function is constant. Note that when β is equal to one, the Weibull distribution is the exponential distribution.

There is a useful relationship between the exponential MTTF, $\theta = 1/\gamma$, and the Weibull parameters λ and β (when the exponential and Weibull MTTFs are equal):

$$\frac{1}{\gamma} = \frac{\Gamma\left(\frac{1}{\beta}\right)}{\beta\lambda} \Rightarrow \lambda = \frac{\gamma\Gamma\left(\frac{1}{\beta}\right)}{\beta} \Rightarrow \lambda = \frac{\Gamma\left(\frac{1}{\beta}\right)}{\beta\theta}. \tag{2.10}$$

This relationship will be useful when comparing the performance of the two distributions.

Let α = confidence level,
T_2 = time to 2nd failure,
$N(t)$ = number of failures to occur in $[0, t]$.

The time, t, needed for $100(\alpha)$ percent confidence, with no more than one failure allowed during testing, required to demonstrate that the system meets the MAV for the MTTF, θ, may be determined, using an adaptation of the test for the exponential distribution, equation (2.5), to the Weibull distribution as follows:

$$1-\alpha \geq P(T_2 > t) = P(N(t) \leq 1) = P(N(t)=0) + P(N(t)=1)$$

$$P(N(t)=0) = e^{-(\lambda t)^\beta}$$

$$P(N(t)=1) = \int_0^t \overline{F}_{T_2}(t-x) f_{T_2}(x)\,dx = \beta\lambda^\beta \int_0^t x^{\beta-1} e^{-[(\lambda t-\lambda x)^\beta + (\lambda x)^\beta]}\,dx$$

$$1-\alpha \geq e^{-(\lambda t)^\beta} + \beta\lambda^\beta \int_0^t x^{\beta-1} e^{-[(\lambda t-\lambda x)^\beta + (\lambda x)^\beta]}\,dx \tag{2.11}$$

Solution of equation (2.11), with α, β, and λ specified gives the appropriate time to test, t. Note, however, that there is no closed form solution for t in this inequality. It must be solved by numerically integrating for a number of t-values.

Implicit in the preceding model, again, is the notion that if the expected value of the time between failures for one system is greater than or equal to the expected value of the time between failures for a second system, then the probability of having J or fewer failures in a renewal process observed for a fixed time for the first system is at least as large as that for the second system given the same fixed time.

Let X = time to 1st failure of system 1, $X \sim \text{Weibull}(\beta, \lambda_1)$,
Y = time to 1st failure of system 2, $Y \sim \text{Weibull}(\beta, \lambda_2)$, $\lambda_1 \le \lambda_2$, $\beta \ge 0$.

Proposition 2: If $E[X] \ge E[Y]$ then $P(N_X(t) \le 1) \ge P(N_Y(t) \le 1)\ \forall\, t$.

Proof: $\lambda_1 \le \lambda_2 \quad \Rightarrow \quad E[X] = \dfrac{\Gamma(1/\beta)}{\beta\lambda_1} \ge E[Y] = \dfrac{\Gamma(1/\beta)}{\beta\lambda_2}$

$$\lambda_1 \le \lambda_2 \quad \Rightarrow \quad \bar{F}_X(t) = e^{-(\lambda_1 t)^\beta} \ \ge\ \bar{F}_Y(t) = e^{-(\lambda_2 t)^\beta} \quad \forall\, t \ge 0, \beta \ge 0$$

Let $N_X(t)$ = the number of renewals that occur in a renewal process having inter arrival distribution $F_X(t)$,
$N_Y(t)$ = the number of renewals that occur in a renewal process having inter arrival distribution $F_Y(t)$.

then $\bar{F}_X(t) \ge \bar{F}_Y(t) \Rightarrow N_X(t) \le_{ST} N_Y(t)\ \ \forall\, t \ge 0$ (Ross, p.257, 1983)

which means that $P(N_X(t) \le 1) \ge P(N_Y(t) \le 1)$. *Q.E.D.*

D. TEST TIME NEEDED TO DEMONSTRATE A THRESHOLD

Suppose COTF is testing a new radar system whose MAV for MTTF is 1000 hours. Suppose further, for the time being, that the radar system's true MTTF is in fact 1000 hours, i.e., $\gamma = 0.001$. Using the assumption of a constant failure rate function, and the thumb rule of 80 percent confidence with no more than one failure, COTF would then test the system for 3000 hours and allow at most one failure to occur. If the system passes this test, COTF will report that, with 80 percent confidence, the radar system MTTF meets or exceeds the MAV.

Now suppose COTF wishes to model this system's behavior with a Weibull distribution. Several choices for the shape parameter, β, are available. Assume that, depending on the platform on which the radar system will be deployed, the shape parameter may vary anywhere between 0.5 and 3.0. The relationship derived in equation (2.10) can now be used to match the MTTF of the exponential distribution, $1/\gamma$, to the MTTF of a Weibull distribution given various shape parameters, β, as seen in Table 1.

TABLE 1
RELATIONSHIP OF EXPONENTIAL FAILURE RATE, γ, TO WEIBULL SCALE PARAMETER, λ, TO MATCH MTTFs FOR DIFFERENT WEIBULL SHAPE PARAMETERS, β

Exponential failure rate, γ	Weibull shape parameter, β	$\Gamma(1/\beta)$	Weibull scale parameter, λ
0.001	0.5	1	0.002
0.001	1	1	0.001
0.001	2	1.7725	0.000886
0.001	3	2.6789	0.000893

Now it is possible to use equation (2.11) to derive the appropriate test times for the preceding values of β as shown in Table 2, again assuming a confidence level of 80 percent, and allowing for at most one failure to occur.[2]

[2]The author solved equation (2.11) using MathCad version 3.1, varying t in one hour increments until the inequality was satisfied.

TABLE 2
TEST TIMES ASSOCIATED WITH VARIOUS WEIBULL SHAPE PARAMETERS, β

β	λ	t (hrs)
0.5	0.002	2950
1	0.001	3000
2	0.000886	2613
3	0.000893	2435

This analysis suggests that, everything else being equal, the assumption of an exponential failure distribution ($\beta = 1$) generates the longest test time to demonstrate that the system meets or exceeds the MAV for MTTF at an 80 percent confidence level. In other words, if it is reasonable to assume that the system under test follows a Weibull time to failure distribution, and if the OTA can identify a reasonable β value for the system (where $\beta \neq 1$), then by modeling the system with a Weibull failure distribution the OTA may be able to reduce the required number of hours needed to demonstrate that the system meets or exceeds the MAV for MTTF for a Weibull distribution with the same shape parameter while maintaining the same level of confidence in its result.

E. USE OF EXPONENTIAL MODEL ON A WEIBULL PROBLEM

Now suppose COTF uses its traditional rule for determining time needed to test the radar system. Earlier it was shown that for an exponential failure distribution with a MAV for MTTF of 1000 hours, it was necessary to test the radar system for 3000 hours, allowing for at most one failure, to obtain 80 percent confidence that the true MTTF met or exceeded the MAV. Suppose it is later discovered that the true underlying failure distribution was actually Weibull with shape parameter $\beta \neq 1$. What can then be said about the actual level of confidence achieved from the test? Is it still 80 percent?

One way to explore this is to hold the test time constant at 3000 hours, and determine the probability of passing the test ($1 - \alpha$) and also the level of confidence (α) achieved when the underlying failure distribution is Weibull. This can be accomplished using equation (2.11) by fixing the time to test, t, at 3000 hours, specifying β and λ, and then solving for α. Results are shown in Table 3.[3]

This analysis seems to suggest that no confidence is lost by assuming an exponential failure distribution when the true underlying failure distribution is Weibull with parameter $\beta > 1$. This makes intuitive sense. As was suggested back in Table 2, for any $\beta > 1$, the required test time is reduced, thus if the system is tested longer than necessary to achieve 80 percent confidence, as occurs in Table 3 when $\beta > 1$, then it naturally follows that the actual confidence level in the result of the test should increase.

TABLE 3
ACTUAL CONFIDENCE LEVELS ACHIEVED AS A FUNCTION OF THE SHAPE PARAMETER, β, WHEN EXPONENTIAL DISTRIBUTION IS USED TO MODEL A WEIBULL (β,λ) PROBLEM

β	λ	t (hrs)	P(pass test) ($1-\alpha$)	Conf. level (α)
0.5	0.002	3000	0.196	0.804
1	0.001	3000	0.200	0.800
2	0.000886	3000	0.097	0.903
3	0.000893	3000	0.028	0.972

Unfortunately, the results of this analysis may not hold under all circumstances. The reason for this is because Propositions 1 and 2 no longer apply. In other words, the fact that $E[X] \geq E[Y]$ no longer implies that $P(N_X(t) \leq 1) \geq P(N_Y(t) \leq 1)$.

Let X = time to 1st failure of system 1, $X \sim$ exponential(γ),
Y = time to 1st failure of system 2, $Y \sim$ Weibull(β, λ).

[3]Results obtained by solving equation (2.11) using MathCad 3.1, varying α until the inequality was met.

Assume: $E[X] = \frac{1}{\gamma} \geq E[Y] = \frac{\Gamma(1/\beta)}{\beta\lambda}$

Unfortunately $\frac{1}{\gamma} \leq \frac{\Gamma(1/\beta)}{\beta\lambda}$ does not imply that

$$\bar{F}_X(t) = e^{-(\gamma t)} \geq \bar{F}_Y(t) = e^{-(\lambda t)^\beta} \quad \forall\, t \geq 0, \beta \geq 0.$$

Therefore, $P(N_X(t) \leq 1) \geq P(N_Y(t) \leq 1)$ cannot be shown using the proof given for Propositions 1 and 2.

Thus, the reader is cautioned not to generalize the result obtained in the previous analysis. It is unclear how the renewal process with exponentially distributed inter arrival times compares to the renewal process with Weibull distributed inter arrival times when their mean arrival times are equated.

F. CONCLUSIONS

The preceding analyses suggest that the OTA can in fact reduce the amount of time needed to test a system for mean time to failure if the system under test follows a Weibull failure distribution with shape parameter $\beta > 1$, and if this shape parameter is known or can be estimated. Unfortunately, the shape parameter β is seldom, if ever, known, and it is also difficult to estimate without prior testing, which tends to make moot the notion of using the Weibull failure distribution to reduce needed test resources.[4] So, from this perspective, one can conclude that there is probably little utility in the OTA attempting to model system reliability using the Weibull distribution as previously suggested.

Moreover, from the perspective of confidence levels, it remains unclear how the use of the test based on the exponential distribution, when applied to a Weibull problem,

[4]The Gaver-Jacobs method for "borrowing information" from previous similar systems may be useful in this regard. Details of the Gaver-Jacobs method is given in the next section.

affects the actual confidence levels achieved. Analysis suggests that, for purposes of making inferences about the mean time to failure, it might be perfectly safe for the OTA to use the exponential failure distribution to model a system that would perhaps more appropriately be modeled using the Weibull distribution, since, as suggested in Tables 2 and 3, modeling a system which has a Weibull failure rate function with the exponential distribution seems to cause the OTA to overestimate the number of hours to test, which in turn seems to provide more confidence in the result of the test. It will be desirable to investigate conditions under which this conservatism property is generally true.

Finally, it should be remembered that OTAs such as COTF tend to perform testing at the system level in which there are numerous competing failure modes. It is well-established in the literature (e.g., NAVORDSYSCOM, p. 3-25, 1971) that as the number of competing failure modes increases, the occurrence of any one failure tends to occur at a constant rate, regardless of the individual underlying failure distributions. So, despite the warning concerning modeling a Weibull system with an exponential time between failure distribution, for purposes of making inferences about the mean time to failure, with larger systems it is not necessarily inappropriate to assume an exponential failure distribution with a constant failure rate function.

III. ON COMBINING DT AND OT FAILURE DATA

A. BACKGROUND AND PROBLEM DESCRIPTION

It has recently been postulated that there may be a useful relationship between Developmental Testing (DT) and Operational Testing (OT) time-between-failure data for many systems. In fact it has been noted anecdotally by some researchers that the OT failure rate tends to be roughly four times the DT failure rate for a given system.[5] Gaver and Jacobs (1994) note that if this relationship could be quantified, then it could be used to anticipate the OT failure rate of a new, similar system based on the DT failure rate from that system.

1. Example Problem Setting

An OT&E analyst has in his possession historical time-between-failure data for both the DT and OT phases of ten projects that are similar, e.g., radar systems. The analyst believes a relationship exists between the DT and OT failure rates such that

$$\lambda_o(i) = K\lambda_d(i) \tag{3.1}$$

where $\lambda_o(i)$ represents the OT failure rate of system i, $\lambda_d(i)$ represents the DT failure rate of system i, and K represents a constant of proportionality[6] and where the times between failures are independent, having an exponential distribution with means $1/\lambda_o(i)$ and $1/\lambda_d(i)$, respectively. The analyst further believes that if he can determine the

---

[5]Interview between D. P. Gaver, Professor of Operations Research, Naval Postgraduate School, and the author, March 1994.

[6]The author recognizes that this "constant" of proportionality is actually a random variable with a distribution of its own. Refer to section E of this chapter for a Bayesian method of estimation of K.

value of K, then he can use this value on an 11th similar system to anticipate the OT failure rate of System 11 given System 11's DT failure data.

B. THE GAVER-JACOBS MODEL

Suppose $T_d(i;j)$ is a random variable representing the time between the (j-1)st and the jth time to failure in DT of system i where failures occur according to a Poisson Process such that

$$P\{T_d(i,j) > t_d\} = e^{-\lambda_d(i)t_d} \quad , \quad t_d \geq 0. \tag{3.2}$$

Suppose further that a similar relationship holds for $T_o(i;j)$, the random variable representing the time between the (j-1)st and the jth time to failure in OT of system i

$$P\{T_o(i,j) > t_o\} = e^{-\lambda_o(i)t_o} \quad , \quad t_o \geq 0. \tag{3.3}$$

Then $N_d(t_d(i))$ is the number of failures to occur during DT for a test of length $t_d(i)$ fixed in advance, and has a Poisson distribution, i.e.:

$$P\{N_d(t_d(i)) = n\} = e^{-\lambda_d(i)t_d(i)} \frac{(\lambda_d(i)t_d(i))^n}{n!}, \; n = 0, 1, ... \tag{3.4}$$

An equivalent distribution holds for OT.

Assume that for each system i, there are $n_d(i)$ DT failures in a test of length $t_d(i)$, and $n_o(i)$ OT failures in a test of length $t_o(i)$. Gaver and Jacobs (p. 3, 1994) have devised a maximum likelihood estimator for the constant of proportionality K as follows:

The likelihood function is

$$L(\underline{\lambda}_d, \underline{\lambda}_o; \text{data}) = \prod_{i=1}^{I} e^{-\lambda_d(i)t_d(i)} \frac{(\lambda_d(i))^{n_d(i)}}{n_d(i)!} e^{-\lambda_o(i)t_o(i)} \frac{(\lambda_o(i))^{n_o(i)}}{n_o(i)!} \tag{3.5}$$

where I denotes the total number of similar systems. Now assume $\lambda_o(i) = K\lambda_d(i)$; then, up to irrelevant constants, the likelihood of K and the development parameters is

$$\begin{aligned} L(K, \lambda_d(i); \text{data}) &= e^{-\sum_i \lambda_d(i)t_d(i) - K\sum_i \lambda_d(i)t_o(i)} \prod_{i=1}^{I} \lambda_d(i)^{n_d(i)} K^{n_o(i)} \lambda_d(i)^{n_o(i)} \\ &= K^{n_o} \prod_{i=1}^{I} e^{-\lambda_d(i)(t_d(i)+Kt_o(i))} \lambda_d(i)^{n_d(i)+n_o(i)} \end{aligned} \tag{3.6}$$

where $n_o = \sum_{i=1}^{I} n_o(i)$, the total number of OT failures across all systems. Focusing on the parameter K above, take the natural logarithms of both sides of equation (3.6) to get

$$\ell(K, \lambda_d(i); \text{data}) = n_o \ln K - \sum_{i=1}^{I} \lambda_d(i)\left(t_d(i) + Kt_o(i)\right) \tag{3.7}$$

then differentiate with respect to K to get

$$\frac{\partial \ell}{\partial K} = \frac{n_o}{K} - \overline{\lambda_d t_o} \tag{3.8}$$

then finally set equation (3.8) equal to zero and solve for K to get the maximum likelihood estimator

$$\hat{K} = \frac{n_o}{\overline{\lambda_d t_o}} \tag{3.9}$$

where $\overline{\lambda_d t_o} = \sum_{i=1}^{I} \lambda_d(i) t_o(i)$, and $t_O(i)$ is the length of time over which system i is subjected to test in OT. Unfortunately $\lambda_d(i)$ is unknown, but may be estimated as follows:

$$\hat{\lambda}_d(i) = \frac{n_d(i)}{t_d(i)}. \tag{3.10}$$

The properties of this estimator for K are as yet unknown and will be investigated shortly. Meanwhile, notice that, assuming $\hat{K}$ is well-behaved, the analyst now has a mechanism for relating past DT failure data on a system to future OT failure data on the same system given historical DT and OT failure data from similar systems. Specifically

$$\hat{\lambda}_o(i) = \hat{K}\hat{\lambda}_d(i). \tag{3.11}$$

C. ANALYSIS OF THE GAVER-JACOBS MODEL

The test of the performance of the estimator, $\hat{K}$, and the follow-on estimate $\hat{\lambda}_o(i)$, was conducted using a Monte Carlo simulation. The idea was to simulate ten "similar" systems (Systems 1 through 10) whose actual OT failure rates were four times their actual DT failure rates. Each system was tested for a length of time, t, such that the mean number of failures in (0, t], λt, was equal to 4.0 for each system.[7] Associated with each system was a random number seed which was used with a Poisson random number generator, written by the author in Turbo Pascal 6.0, the code for which is included in Appendix B. A summary of the initial data is given in Table 4.

1. Estimation of K.

Next, 1000 separate observations of the number of OT failures and DT failures observed over the appropriate test time for each system were generated using the random seed indicated in Table 4. Then, $\hat{K}$ was derived using equations (3.9) and (3.10) for each of the 1000 repetitions yielding 1000 separate observations of $\hat{K}$. Appropriate statistics were gathered on $\hat{K}$. Table 5 summarizes these statistics and Figure 1 shows a histogram of the observations of $\hat{K}$.

[7]This number was chosen arbitrarily, but was kept constant to simulate consistent choice of test duration.

TABLE 4
INITIAL DATA ON TEN "SIMILAR" SYSTEMS IN WHICH
ACTUAL K = 4.0 AND ACTUAL POISSON FAILURE RATE $\lambda t = 4.0$

System number	Actual DT failure rate	DT test time (hours)	DT Random seed	Actual OT failure rate	OT test time (hours)	OT Random seed
1	0.0002	20,000	10	0.0008	5,000	110
2	0.0004	10,000	20	0.0016	2,500	120
3	0.0006	6,666.67	30	0.0024	1,666.67	130
4	0.0008	5,000	40	0.0032	1,250	140
5	0.001	4,000	50	0.004	1,000	150
6	0.002	2,000	60	0.008	500	160
7	0.004	1,000	70	0.016	250	170
8	0.006	666.67	80	0.024	166.67	180
9	0.008	500	90	0.032	125	190
10	0.01	400	100	0.04	100	200

TABLE 5
SUMMARY STATISTICS ON 1000
OBSERVATIONS OF $\hat{K}$

Mean	4.040
Std Error of the Mean	0.030
Median	4.000
Std Dev of the simulated values of K	0.939
Variance of the simulated values of K	0.882
Kurtosis	0.677
Skewness	0.675
Range	5.779
Minimum	1.913
Maximum	7.692
Sum	4039.955
Count	1000

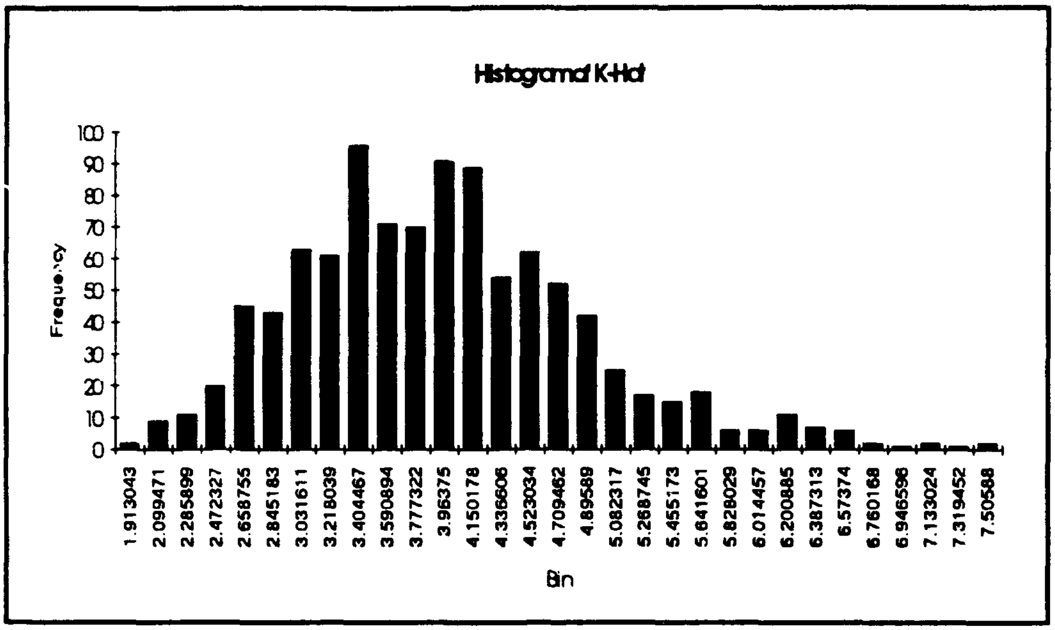

Figure 1 - Histogram of 1000 observations of estimate of K.

2. Estimation of the OT failure rate.

While information about the sampling distribution of $\hat{K}$ is interesting, the real focus of this study is on how well the analyst can predict the OT failure rate. Three methods for estimating the OT failure rate are being considered in this study:

1. Prediction using the relationship $\hat{\lambda}_o(i) = \hat{K}\hat{\lambda}_d(i)$;
2. Direct observation of the OT failure rate from OT data; and
3. Weighted average of the two previous methods.

In each of the three methods, the following are calculated:

1. The estimate for the OT failure rate;
2. The variance of the estimate, derived appropriately from each method;
3. The standard error (square root of the variance) of the estimate; and
4. The mean-squared error, which is the sum of the square of the average of the biases and the average of the variances over all replications.

Figure 2, below, describes one replication of the OT failure rate estimation simulation model.

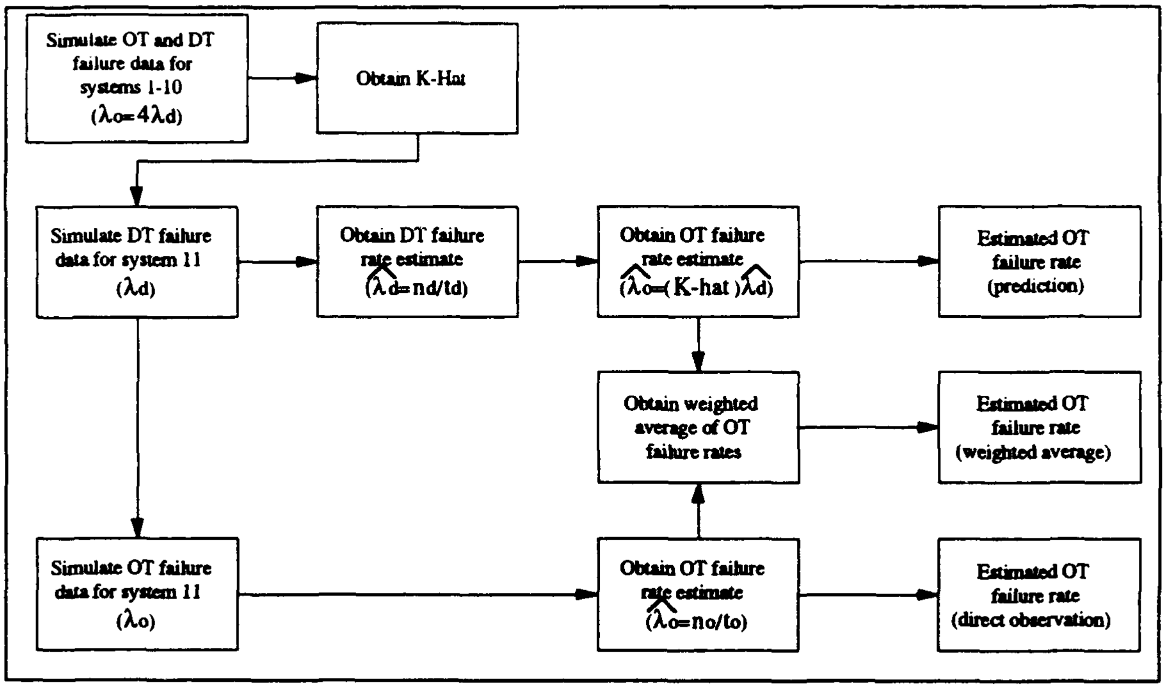

Figure 2 - One replication of the OT failure rate estimation simulation model.

a. Prediction using the estimator for K.

The analyst's prediction of the future OT failure rate for an 11th system, using the estimator for K, would be

$$\hat{\lambda}_o(11) = \hat{K}\hat{\lambda}_d(11). \tag{3.12}$$

The associated variance would be computed as follows:

$$\begin{aligned}
\mathrm{Var}[\hat{\lambda}_o(11)] &= \mathrm{Var}[\hat{K}\hat{\lambda}_d(11)] \\
&= E[\,(\hat{K}\hat{\lambda}_d(11) - E[\hat{K}]E[\hat{\lambda}_d(11)]\,)^2] \\
&= E[\hat{K}^2\hat{\lambda}_d(11)^2 - 2\hat{K}\hat{\lambda}_d(11)E[\hat{K}]E[\hat{\lambda}_d(11)] + E[\hat{K}]^2E[\hat{\lambda}_d(11)]^2] \\
&= E[\hat{K}^2]E[\hat{\lambda}_d(11)^2] - E[\hat{K}]^2E[\hat{\lambda}_d(11)]^2] \\
&= (\mathrm{Var}[\hat{K}] + E[\hat{K}]^2)(\mathrm{Var}[\hat{\lambda}_d(11)] + E[\hat{\lambda}_d(11)]^2) - E[\hat{K}]^2E[\hat{\lambda}_d(11)]^2.
\end{aligned}$$

simplifying gives:

$$\mathrm{Var}[\hat{\lambda}_o(11)] = \mathrm{Var}[\hat{K}]\mathrm{Var}[\hat{\lambda}_d(11)] + \mathrm{Var}[\hat{K}]\mathrm{E}[\hat{\lambda}_d(11)]^2\,) + \mathrm{Var}[\hat{\lambda}_d(11)]\mathrm{E}[\hat{K}]^2 \;. \quad (3.13)$$

Expressions are needed for $\mathrm{Var}[\hat{K}]$ and $\mathrm{Var}[\hat{\lambda}_d(11)]$. An approximation to $\mathrm{Var}[\hat{K}]$ can be derived through maximum likelihood estimation using observed Fisher information as follows (Cox and Hinkley, p.297, 1974):

$$\mathrm{Var}[\hat{K}] \approx -\frac{1}{\left.\frac{\partial^2 \ell}{\partial K^2}\right|_{K=\hat{K}}}. \qquad (3.14)$$

Taking the second derivative with respect to $\hat{K}$ of equation (3.7) gives

$$\frac{\partial^2 \ell}{\partial \hat{K}^2} = -\frac{n_o}{\hat{K}^2}, \qquad (3.15)$$

Thus

$$\mathrm{Var}[\hat{K}] \approx \frac{\hat{K}^2}{n_o} = \frac{\left(n_o \big/ \overline{\lambda_d t_o}\right)^2}{n_o} = \frac{n_o}{\left(\overline{\lambda_d t_o}\right)^2} \approx \frac{n_o}{\left(\sum \frac{n_d(i)}{t_d(i)} t_o(i)\right)^2}. \qquad (3.16)$$

$\mathrm{Var}[\hat{\lambda}_d(11)]$ can be derived as follows by noting that $\hat{\lambda}_d(i)$ is simply a Poisson random variable, $N_d(i)$, multiplied by a constant, $\frac{1}{t_d(i)}$:

$$\begin{aligned} \mathrm{Var}[\hat{\lambda}_d(i)] &= \mathrm{Var}\left(\frac{N_d(i)}{t_d(i)}\right) = \frac{1}{(t_d(i))^2}\mathrm{Var}[N_d(i)] \\ &= \frac{1}{(t_d(i))^2}\lambda_d(i) t_d(i) \approx \frac{n_d(i)}{(t_d(i))^2} \end{aligned} \qquad (3.17)$$

The standard error of $\hat{\lambda}_d(11)$ is then simply the square root of $\mathrm{Var}[\hat{\lambda}_d(11)]$.

An example prediction can now be performed. Suppose, using data from the first simulation run, equation (3.9) gives $\hat{K}$ = 4.293 and equation (3.10) gives $\hat{\lambda}_d(11)$ = 0.0003. Then

$$\hat{\lambda}_o(11) = \hat{K}\hat{\lambda}_d(11) = (4.293)(0.0003) = 0.00129.$$

Note that per the original assumption, K = 4.0, so we would expect a system whose DT failure rate is 0.0003 to have an OT failure rate of 0.0012. So, at first glance this appears to be a relatively good estimate of the true situation.

The variance and standard error of $\hat{\lambda}_o(11)$ is solved using equation (3.13) as follows:

$$E[\hat{K}] \approx \hat{K} = 4.293$$
$$E[\hat{\lambda}_d(11)] \approx \hat{\lambda}_d(11) = 0.0003$$
$$\mathrm{Var}[\hat{K}] \approx \frac{n_o}{\left(\sum_{i=1}^{10} \frac{n_d(i)}{t_d(i)} t_o(i)\right)^2} \approx \frac{n_o}{\left(\overline{\lambda_d t_o}\right)^2} = \frac{44}{(10.25)^2} = 0.4188$$
$$\mathrm{Var}[\hat{\lambda}_d(11)] \approx \frac{n_d(11)}{(t_d(11))^2} = \frac{6}{(20{,}000)^2} = 1.5\times10^{-8}$$

finally

$$\mathrm{Var}[\hat{\lambda}_o(11)] = \mathrm{Var}[\hat{K}\hat{\lambda}_d(11)] = (0.4188)(1.5\times10^{-8}) + (1.5\times10^{-8})(4.293)^2 + (0.4188)(0.0003)^2 = 32\times10^{-8}.$$

And the standard error of $\hat{\lambda}_o(11)$ is $\sqrt{32\times10^{-8}}$ = 5.66×10^{-4}. Table 6 summarizes the results of this prediction.

TABLE 6
SUMMARY OF RESULTS OF PREDICTION OF OT FAILURE RATE
ON SYSTEM 11 USING ESTIMATE OF K BASED ON FIRST REPLICATION

Actual Failure Rate $\lambda_o(11)$	0.0012
Estimated Failure Rate $\hat{\lambda}_o(11)$	0.00129
Variance$[\hat{\lambda}_o(11)]$	32×10^{-8}
Standard Error$[\hat{\lambda}_o(11)]$	5.66×10^{-4}
$\hat{\lambda}_o(11) \pm 1$ Standard Error	(0.0007, 0.0019)
$\hat{\lambda}_o(11) \pm 2$ Standard Errors	(0.0002, 0.0024)

b. Direct observation from OT failure data.

Of course, the analyst could, if he desired, come up with a value for $\hat{\lambda}_o(11)$ by observing the empirical OT failure rate as usual. In this situation, the analyst would note the number of failures which occur, $n_o(11)$, during the time on test, $t_o(11)$, and derive the OT failure rate as follows:

$$\hat{\lambda}'_o(11) = \frac{n_o(11)}{t_o(11)}. \tag{3.18}$$

For example, suppose that System 11 is operationally tested for a period of 5,000 hours and experiences 6 failures during that time. Then the OT failure rate observed through operational testing would be

$$\hat{\lambda}'_o(11) = \frac{6}{5000} = 0.0012.$$

Using the same reasoning as in equation (3.17), the variance of the observed value is calculated as

$$\mathrm{Var}[\hat{\lambda}_o'(11)] = \mathrm{Var}\left[\frac{N_o(11)}{t_o(11)}\right] = \left(\frac{1}{t_o(11)}\right)^2 \mathrm{Var}(N_o(11))$$

$$= \left(\frac{1}{t_o(11)}\right)^2 \lambda_o(11)t_o(11) \approx \left(\frac{1}{t_o(11)}\right)^2 n_o(11)$$

$$= \frac{6}{(5000)^2} = 24 \times 10^{-8}\,.$$

And the standard error of $\hat{\lambda}_o'(11)$ is $\sqrt{24 \times 10^{-8}} = 4.9 \times 10^{-4}$. Table 7 summarizes the results of these calculations.

TABLE 7
SUMMARY OF RESULTS OF DIRECT OBSERVATION OF OT FAILURE RATE ON SYSTEM 11 FROM OT FAILURE DATA

Actual Failure Rate $\lambda_o(11)$	0.0012
Estimated Failure Rate $\hat{\lambda}_o'(11)$	0.0012
Variance$[\hat{\lambda}_o'(11)]$	24×10^{-8}
Standard Error$[\hat{\lambda}_o'(11)]$	4.9×10^{-4}
$\hat{\lambda}_o'(11) \pm 1$ Standard Error	(0.0007, 0.0017)
$\hat{\lambda}_o'(11) \pm 2$ Standard Errors	(0.0002, 0.0022)

Unfortunately, this method would not save the analyst any time since he would have to run the entire operational test procedure to derive a meaningful number. Moreover, he would not be taking advantage of the available DT failure data.

c. Weighted average of predicted value and direct observation.

Suppose now that the analyst runs the entire operational test procedure as above, but also desires to use the available DT failure data in an effort to improve his estimate for the OT failure rate. Assuming the analyst has the data from the 10 other previous systems with which to derive the prediction, $\hat{\lambda}_o(11) = \hat{K}\hat{\lambda}_d(11)$, he can then calculate a weighted average of the two estimates.

An estimator for the mean of a linear combination of two random variables, X and Y, with common mean *m* and different variances, σ_X^2 and σ_Y^2 can be constructed as follows:

$$\text{Let} \quad \hat{m} = aX + bY, \qquad 0 \le a,b \le 1, \qquad a + b = 1.$$

$$\text{Then} \quad E[\hat{m}] = am + bm = m(a + b) = m(1) = m,$$

$$\begin{aligned}
\ell(a) &= E[(\hat{m} - m)^2] = E[(aX + (1-a)Y - m)^2] \\
&= E[(a(X - m) + (1 - a)(Y - m))^2] \\
&= E[a^2(X-m)^2 + 2a(1-a)(X-m)(Y-m) + (1-a)^2(Y-m)^2] \\
&= a^2\text{Var}[X] + (1 - a)^2\text{Var}[Y].
\end{aligned}$$

Differentiating with respect to a and setting equal to zero gives

$$\frac{\partial \ell(a)}{\partial a} = 2a\text{Var}[X] - 2(1 - a)\text{Var}[Y] = 0$$

which yields

$$a[\text{Var}[X] + \text{Var}[Y]] = \text{Var}[Y]$$

or

$$a = \frac{\text{Var}[Y]}{\text{Var}[X] + \text{Var}[Y]} = \frac{1/\sigma_X^2}{1/\sigma_X^2 + 1/\sigma_Y^2}.$$

Similarly

$$b = \frac{\text{Var}[X]}{\text{Var}[X] + \text{Var}[Y]} = \frac{1/\sigma_Y^2}{1/\sigma_X^2 + 1/\sigma_Y^2}.$$

Substitution then yields

$$\hat{m} = \frac{X/\sigma_X^2 + Y/\sigma_Y^2}{1/\sigma_X^2 + 1/\sigma_Y^2}. \tag{3.19}$$

Therefore an estimator for the OT failure rate, $\lambda_o(i)$, can be constructed as follows:

$$\hat{\hat{\lambda}}_o(i) = \frac{\dfrac{\hat{K}\hat{\lambda}_d(i)}{\hat{V}ar[\hat{K}\hat{\lambda}_d(i)]} + \dfrac{\hat{\lambda}'_o(i)}{Var[\hat{\lambda}'_o(i)]}}{\dfrac{1}{\hat{V}ar[\hat{K}\hat{\lambda}_d(i)]} + \dfrac{1}{Var[\hat{\lambda}'_o(i)]}} \tag{3.20}$$

where $\hat{K}\hat{\lambda}_d(i)$ represents the predicted value of the OT failure rate and $\hat{\lambda}'_o(i)$ denotes the empirical OT failure rate observed during operational testing.

For example, the values obtained previously for $\hat{K}\hat{\lambda}_d(11)$, the predicted value, and $\hat{\lambda}'_o(11)$, the observed value, were 0.00129 and 0.0012 respectively. It was shown earlier that the estimated variance for the predicted value was $\hat{V}ar[\hat{K}\hat{\lambda}_d(11)] = 32 \times 10^{-8}$, and the variance of the observed value was $Var[\hat{\lambda}'_o(11)] = 24 \times 10^{-8}$. Inserting these numbers in equation (3.20) gives

$$\hat{\hat{\lambda}}_o(11) = \frac{\dfrac{12.9\times10^{-4}}{32\times10^{-8}} + \dfrac{12\times10^{-4}}{24\times10^{-8}}}{\dfrac{1}{32\times10^{-8}} + \dfrac{1}{24\times10^{-8}}} = 12.39\times10^{-4}.$$

The variance associated with this estimate can be constructed as follows:

Let $\quad Var[\hat{m}] = Var[aX + bY] = a^2 Var[X] + b^2 Var[Y].$

Substitution yields

$$\sigma_m^2 = \frac{\left(\frac{1}{\sigma_X^2}\right)^2 \sigma_X^2 + \left(\frac{1}{\sigma_Y^2}\right)^2 \sigma_Y^2}{\left(\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2}\right)^2} = \frac{1}{\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2}}. \tag{3.21}$$

Therefore, the estimated variance of $\hat{\hat{\lambda}}_o(11)$ is computed as follows:

$$\hat{\mathrm{V}}\mathrm{ar}[\hat{\hat{\lambda}}_o(11)] = \frac{1}{\frac{1}{\hat{\mathrm{V}}\mathrm{ar}[\hat{K}\hat{\lambda}_d(11)]} + \frac{1}{\mathrm{Var}[\hat{\lambda}_o'(11)]}} = \frac{1}{\frac{1}{32\times10^{-8}} + \frac{1}{24\times10^{-8}}} = 13.7\times10^{-8}.$$

And the standard error of $\hat{\hat{\lambda}}_o(11)$ is $\sqrt{13.7\times10^{-8}} = 3.7 \times 10^{-4}$. Table 8 summarizes the results of these calculations.

TABLE 8
SUMMARY OF RESULTS OF WEIGHTED AVERAGE OF
PREDICTED VALUE AND DIRECT OBSERVATION OF OT FAILURE RATE

Actual Failure Rate $\lambda_o(11)$	0.0012
Estimated Failure Rate $\hat{\hat{\lambda}}_o(11)$	0.00124
Variance[$\hat{\hat{\lambda}}_o(11)$]	13.7×10^{-8}
Standard Error[$\hat{\hat{\lambda}}_o(11)$]	3.7×10^{-4}
$\hat{\hat{\lambda}}_o(11) \pm 1$ Standard Error	(0.0009, 0.0016)
$\hat{\hat{\lambda}}_o(11) \pm 2$ Standard Errors	(0.0005, 0.0020)

3. Comparison of the three methods for estimating the OT failure rate.

Table 9 summarizes the results obtained for each of the three methods of deriving the OT failure rate for one replication of the simulation. They are arguably quite similar. So to answer the question of which method is best, further analysis is required.

TABLE 9
COMPARISON OF THE THREE METHODS FOR OBTAINING
THE OT FAILURE RATE

Statistic	Prediction Method	Direct Observation Method	Weighted Average Method
Actual Value	0.0012	0.0012	0.0012
Estimated Value	0.00129	0.0012	0.00124
Variance	32×10^{-8}	24×10^{-8}	13.7×10^{-8}
Standard Error	5.66×10^{-4}	4.9×10^{-4}	3.7×10^{-4}
Estimate ± 1 S.E.	(0.0007, 0.0019)	(0.0007, 0.0017)	(0.0009, 0.0016)
Estimate ± 2 S.E.s	(0.0002, 0.0024)	(0.0002, 0.0022)	(0.0005, 0.0020)

Since the results of the three methods discussed previously were so similar, a look at the asymptotic behavior of the three methods is in order. To facilitate this analysis, four new "similar" systems were created, systems 11, 12, 13 and 14, each with failure rates and test times of similar construction to the previous 10 systems, i.e., $\lambda_o(i) = K\lambda_d(i)$, where K is equal to 4.0 and λt is also equal to 4.0. Each system also had its own random seeds for both DT and OT failures. Table 10 summarizes the four new systems.

TABLE 10
INITIAL DATA ON FOUR NEW "SIMILAR" SYSTEMS IN WHICH
ACTUAL K = 4.0 AND ACTUAL POISSON FAILURE RATES $\lambda t = 4.0$

System number	Actual DT failure rate	DT test time (hours)	DT random seed	Actual OT failure rate	OT test time (hours)	OT random seed
11	0.0003	13,333.33	12	0.0012	3,333.33	112
12	0.0005	8,000	22	0.002	2,000	122
13	0.0007	5,714.29	32	0.0028	1,428.57	132
14	0.0009	4,444.44	42	0.0036	1,111.11	142

One thousand repetitions of a simulated OT and DT were run on each of the four new systems, and then the exact same analysis as that done previously in section III.C.2 was performed for each replication. Grand estimates appear in Tables 11 through 14 in the column entitled "Average value (n = 1000)" and were produced by averaging the estimates of the following from all 1000 simulated observations, or sub-estimates:

1. The estimated OT failure rate derived from prediction, direct observation and weighted average;
2. The estimated variance of the estimated OT failure rate derived from prediction, direct observation and weighted average; and
3. The estimated standard error of the estimated OT failure rate derived from prediction, direct observation and weighted average.

Then, in order to get an idea of the sample distribution of each of the above super-estimates, the sample variance, sample standard deviation and coefficient of variation[8] were computed for the 1000 sub-estimates for each system. Finally, the estimated mean-squared error, which is the sum of the square of the average of the biases with the average of the variances over all replications, i.e.,

$$\hat{MSF} = (\text{Average of the biases})^2 + (\text{Average of the variances}),$$

was calculated for each prediction method and placed in the last column of each table. Figure 3 describes the situation.

[8]The coefficient of variation is simply the sample standard deviation divided by the sample average, and serves as a measure of the spread of the sample.

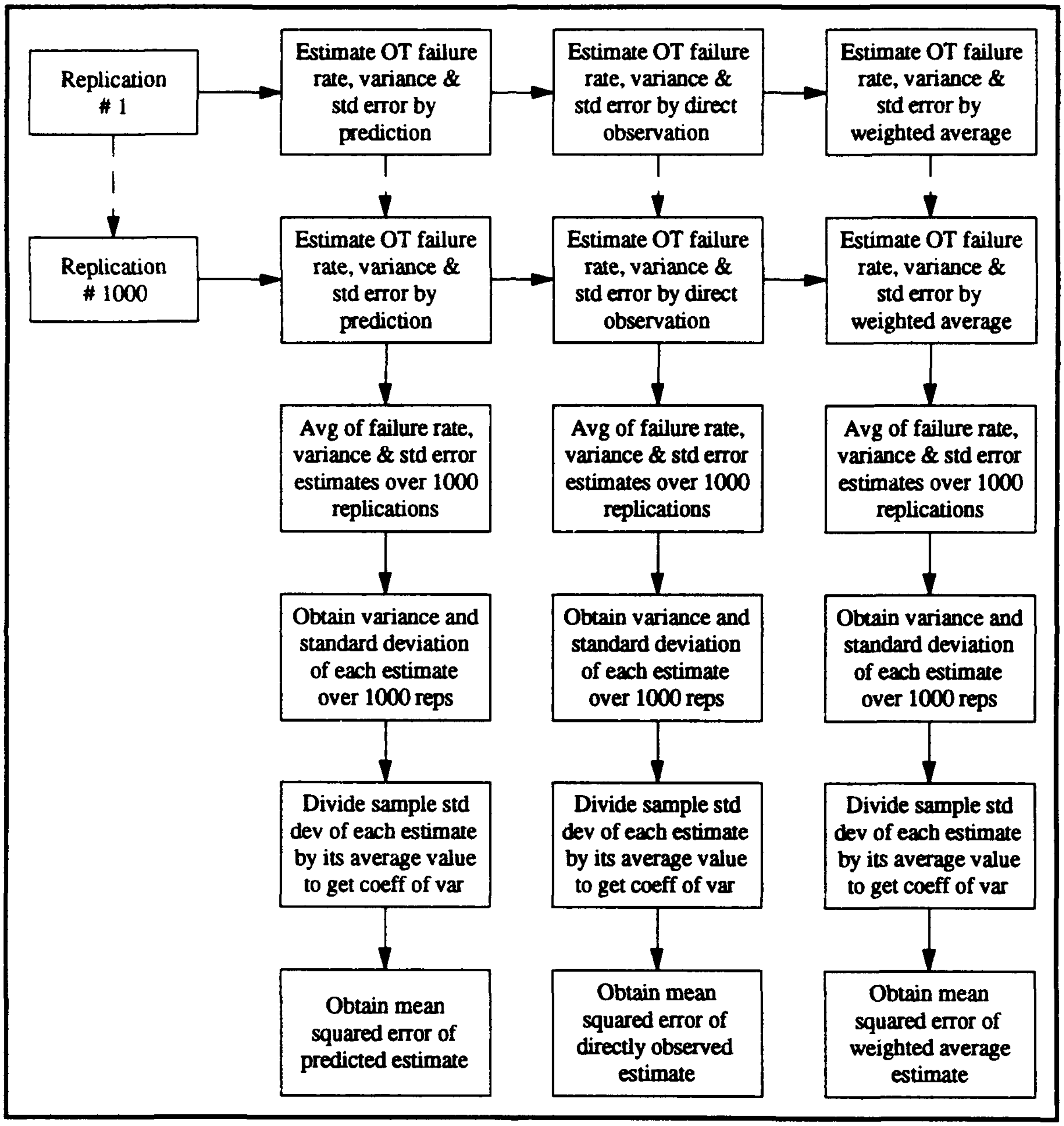

Figure 3 - Methodology for comparing different ways of estimating OT failure rate.

a. New System 11.

The statistics on new System 11 are summarized in Table 11 and a graphical summary is shown in Figure 4. Recall that actual $\lambda_o(11) = 0.0012$.

TABLE 11
NEW SYSTEM 11 - ESTIMATED OT FAILURE RATES AND OTHER RELEVANT STATISTICS (ACTUAL $\lambda_0(11) = 0.0012$)

	Sub-estimates	Average value (n=1000)	Sample Variance (n=1000)	Sample Std Deviation (n=1000)	Coeff of Variation	Mean Squared Error
Prediction	$\hat{\lambda}_o(11)$	0.0012	4.62×10^{-7}	0.00068	57%	4.46×10^{-7}
	$V[\hat{\lambda}_o(11)]$	4.46×10^{-7}	1.17×10^{-13}	3.43×10^{-7}	77%	
	$SE[\hat{\lambda}_o(11)]$	0.00062	5.54×10^{-8}	0.00024	39%	
Direct Observation	$\hat{\lambda}'_o(11)$	0.0012	3.73×10^{-7}	0.00061	51%	3.6×10^{-7}
	$V[\hat{\lambda}'_o(11)]$	3.6×10^{-7}	3.36×10^{-14}	1.83×10^{-7}	51%	
	$SE[\hat{\lambda}'_o(11)]$	0.00058	2.87×10^{-7}	0.00017	29%	
Weighted Average	$\hat{\hat{\lambda}}_o(11)$	0.0010	2.23×10^{-7}	0.00047	47%	2.07×10^{-7}
	$V[\hat{\hat{\lambda}}_o(11)]$	1.67×10^{-7}	6.96×10^{-15}	8.34×10^{-8}	50%	
	$SE[\hat{\hat{\lambda}}_o(11)]$	0.00039	1.4×10^{-8}	0.00012	31%	

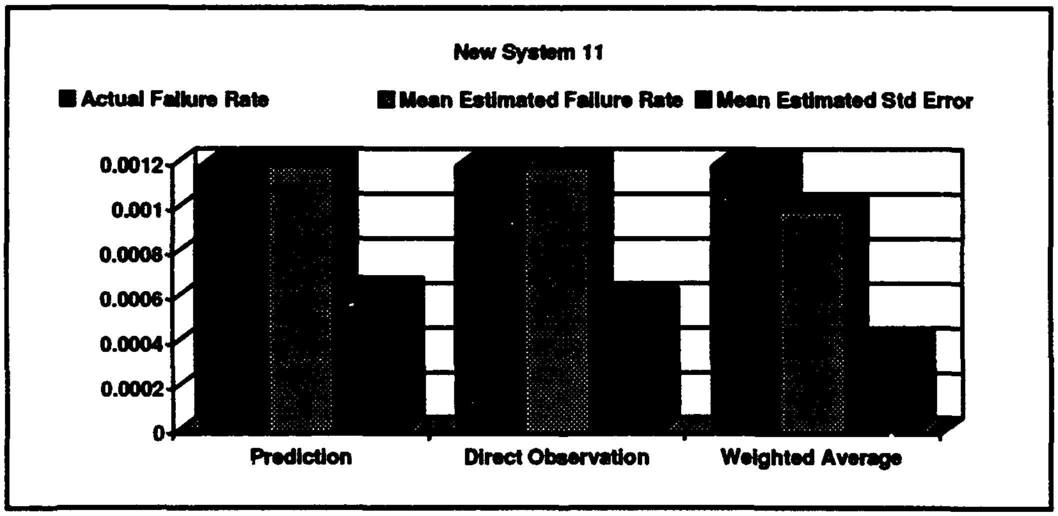

Figure 4 - Graphical summary of System 11 estimates.

b. New System 12.

The statistics on new System 12 are summarized in Table 12 and a graphical summary is shown in Figure 5. Recall that actual $\lambda_o(12) = 0.002$.

TABLE 12
NEW SYSTEM 12 - ESTIMATED OT FAILURE RATES AND OTHER RELEVANT STATISTICS (ACTUAL $\lambda_0(12) = 0.002$)

	Sub-estimates	Average value (n=1000)	Sample Variance (n=1000)	Sample Std Deviation (n=1000)	Coeff of Variation	Mean Squared Error
Prediction	$\hat{\lambda}_o(12)$	0.0020	1.17×10^{-6}	0.0011	55%	1.23×10^{-6}
	$V[\hat{\lambda}_o(12)]$	1.23×10^{-6}	7.51×10^{-13}	8.66×10^{-7}	70%	
	$SE[\hat{\lambda}_o(12)]$	0.0010	1.46×10^{-7}	0.00038	38%	
Direct Observation	$\hat{\lambda}'_o(12)$	0.0020	9.51×10^{-7}	0.00098	49%	9.99×10^{-7}
	$V[\hat{\lambda}'_o(12)]$	9.99×10^{-7}	2.38×10^{-13}	4.88×10^{-7}	49%	
	$SE[\hat{\lambda}'_o(12)]$	0.00096	7.42×10^{-8}	0.00027	28%	
Weighted Average	$\hat{\hat{\lambda}}_o(12)$	0.0017	5.91×10^{-7}	0.00077	45%	5.49×10^{-7}
	$V[\hat{\hat{\lambda}}_o(12)]$	4.59×10^{-7}	5.13×10^{-14}	2.26×10^{-7}	49%	
	$SE[\hat{\hat{\lambda}}_o(12)]$	0.00065	3.71×10^{-8}	0.00019	29%	

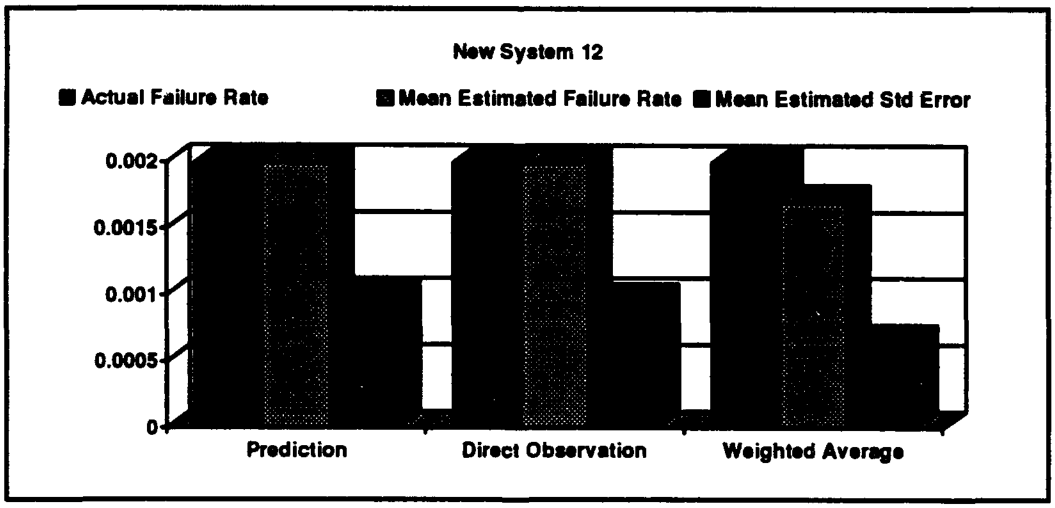

Figure 5 - Graphical summary of System 12 estimates.

c. New System 13.

The statistics on new System 13 are summarized in Table 13 and a graphical summary is shown in Figure 6. Recall that actual $\lambda_o(13) = 0.0028$.

TABLE 13
NEW SYSTEM 13 - ESTIMATED OT FAILURE RATES AND OTHER RELEVANT STATISTICS (ACTUAL $\lambda_o(13) = 0.0028$)

	Sub-estimates	Average value (n=1000)	Sample Variance (n=1000)	Sample Std Deviation (n=1000)	Coeff of Variation	Mean Squared Error
Prediction	$\hat{\lambda}_o(13)$	0.0028	2.33×10^{-6}	0.0015	54%	2.41×10^{-6}
	$V[\hat{\lambda}_o(13)]$	2.41×10^{-6}	3.14×10^{-12}	1.77×10^{-6}	73%	
	$SE[\hat{\lambda}_o(13)]$	0.0015	2.96×10^{-7}	0.00054	36%	
Direct Observation	$\hat{\lambda}'_o(13)$	0.0028	2.00×10^{-6}	0.0014	50%	1.96×10^{-6}
	$V[\hat{\lambda}'_o(13)]$	1.96×10^{-6}	9.82×10^{-13}	9.91×10^{-7}	51%	
	$SE[\hat{\lambda}'_o(13)]$	0.0013	1.52×10^{-7}	0.00039	30%	
Weighted Average	$\hat{\hat{\lambda}}_o(13)$	0.0024	1.24×10^{-6}	0.0011	46%	1.06×10^{-6}
	$V[\hat{\hat{\lambda}}_o(13)]$	9.04×10^{-7}	2.11×10^{-13}	4.59×10^{-7}	51%	
	$SE[\hat{\hat{\lambda}}_o(13)]$	0.00091	7.79×10^{-8}	0.00028	31%	

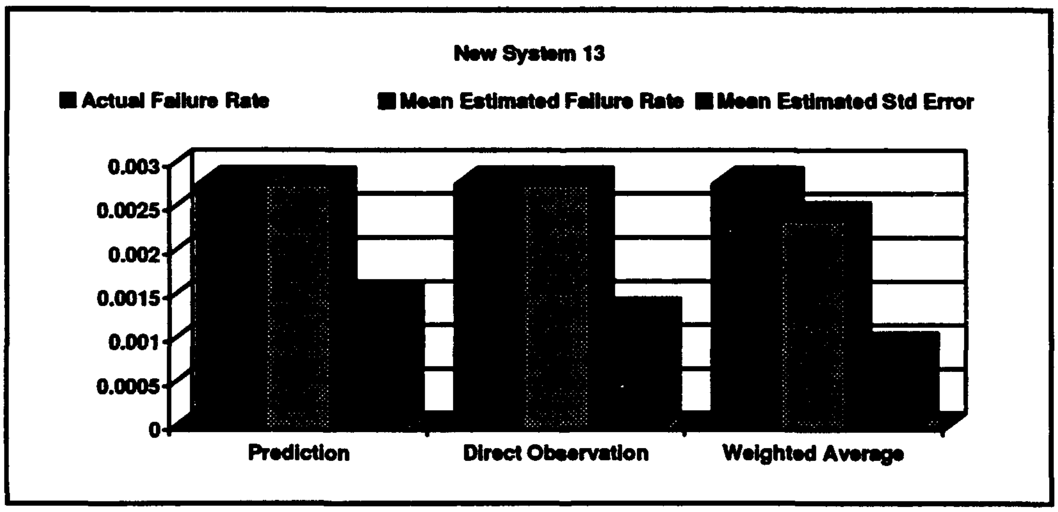

Figure 6 - Graphical summary of System 13 estimates.

d. New System 14.

The statistics on new System 14 are summarized in Table 14 and a graphical summary is shown in Figure 7. Recall that actual $\lambda_o(14) = 0.0036$.

TABLE 14
NEW SYSTEM 14 - ESTIMATED OT FAILURE RATES AND OTHER RELEVANT STATISTICS (ACTUAL $\lambda_0(14) = 0.0036$)

	Sub-estimates	Average value (n=1000)	Sample Variance (n=1000)	Sample Std Deviation (n=1000)	Coeff of Variation	Mean Squared Error
Prediction	$\hat{\lambda}_o(14)$	0.0036	4.07×10^{-6}	0.0020	56%	3.93×10^{-6}
	$V[\hat{\lambda}_o(14)]$	3.93×10^{-6}	8.90×10^{-12}	2.98×10^{-6}	76%	
	$SE[\hat{\lambda}_o(14)]$	0.0018	5.18×10^{-7}	0.00072	40%	
Direct Observation	$\hat{\lambda}'_o(14)$	0.0036	3.20×10^{-6}	0.0018	50%	3.21×10^{-6}
	$V[\hat{\lambda}'_o(14)]$	3.21×10^{-6}	2.59×10^{-12}	1.61×10^{-6}	50%	
	$SE[\hat{\lambda}'_o(14)]$	0.0017	2.45×10^{-7}	0.00050	29%	
Weighted Average	$\hat{\hat{\lambda}}_o(14)$	0.0031	2.05×10^{-6}	0.0014	45%	1.72×10^{-6}
	$V[\hat{\hat{\lambda}}_o(14)]$	1.47×10^{-6}	5.76×10^{-13}	7.59×10^{-7}	52%	
	$SE[\hat{\hat{\lambda}}_o(14)]$	0.0012	1.31×10^{-7}	0.00036	30%	

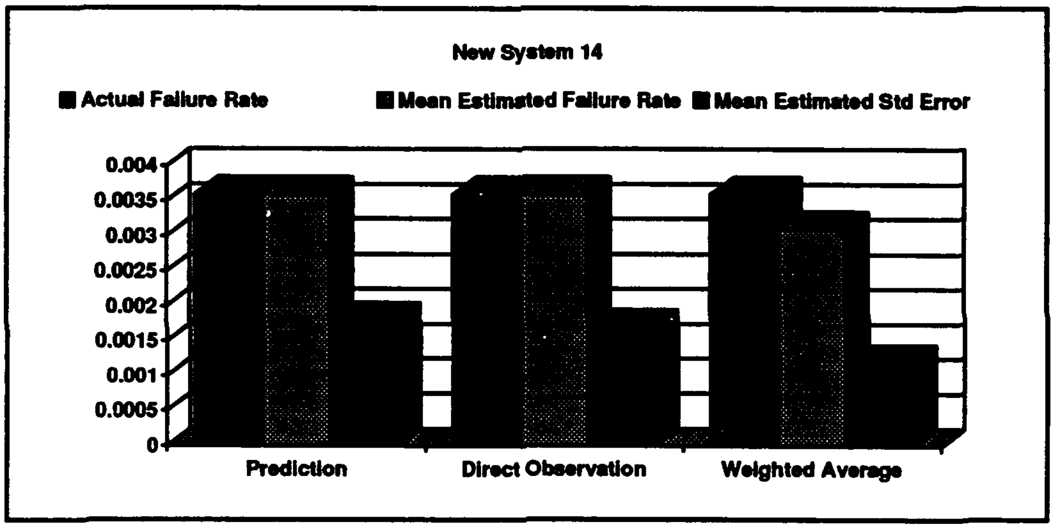

Figure 7 - Graphical summary of System 14 estimates.

D. RESULTS OF THE GAVER-JACOBS MODEL ANALYSIS

Simulation results indicate that the predicted value method and the observed value method are unbiased and provide good estimates of the OT failure rate. The weighted average method, although it gives a more precise estimate because of its smaller mean squared error, is biased in that it tends to underestimate the actual failure rate. The observed value method gives slightly better precision than the predicted value method, since the mean squared error obtained with the observed value method tends to be slightly less than that associated with the predicted value method. Notice, however, that except for the very small price paid in terms of precision, the estimate obtained using the Gaver-Jacobs predicted value method seems to estimate the OT failure rate nearly as well as direct observation of the OT failure rate. This is significant since the Gaver-Jacobs estimate can be used prior to the start of operational testing (provided the constant-K model is nearly correct).

The bottom line is that if the relationship between the DT failure rate and the OT failure rate proposed earlier does in fact exist, then the Gaver-Jacobs method of estimating the OT failure rate of a new system, $\hat{\lambda}_o(i) = \hat{K}\hat{\lambda}_d(i)$, given that system's DT failure rate as well as DT and OT data on previous similar systems, seems to provide a reasonable estimate of the OT failure rate of the new system which performs nearly as well as direct observation of the OT failure rate through operational testing. Moreover, the Gaver-Jacobs estimate of the OT failure rate can be computed prior to the start of operational testing.

E. A HIERARCHICAL BAYES MODEL FOR ESTIMATION OF K

It is idealistic, in the previous model, to assume that all systems have exactly the same value for K.[9] Rather, assume that each system i has its own K value, K_i, and that it is selected independently from a prior distribution. The gamma distribution, for example, turns out to be conjugate, so integrals can be explicitly computed. Its density function is:

$$q_K(K) = e^{-\rho K} \frac{(\rho K)^{\delta-1}}{\Gamma(\delta)} \rho. \tag{3.22}$$

From equations (3.5) and (3.6) it can be seen that the likelihood of K_i, up to irrelevant constants, is

$$L_i(K_i; \lambda_d(i); i-\text{data}) = e^{-K_i \lambda_d(i) t_o(i)} K_i^{n_o(i)} \left[e^{-\lambda_d(i) t_d(i)} (\lambda_d(i))^{n_d(i)+n_o(i)} \right]. \tag{3.23}$$

To obtain the marginal posterior density of K_i, equation (3.22) is re-written, up to irrelevant constants, as follows:

$$q_{K_i}(K_i; i-\text{data}) = e^{-(\rho+\lambda_d(i)t_o(i))K_i} \frac{((\rho+\lambda_d(i)t_o(i))K_i)^{\delta+n_o(i)-1}}{\Gamma(\delta+n_o(i))} (\rho+\lambda_d(i)t_o(i)). \tag{3.24}$$

If the parameters ρ, δ and $\lambda_d(i)$ were known, and the data available, i.e., $n_o(i)$ and $t_o(i)$, then one could compute the conditional expected value of K_i and use this as a point estimate. However there are other, more interesting ways to use the posterior. First, it is

[9]The following explanation for the hierarchical Bayes model was developed by D. P. Gaver, Professor of Operations Research, Naval Postgraduate School, in handwritten notes to the author, July 1994.

necessary to evaluate the various parameters ρ, δ and $\lambda_d(i)$. To do this, the "empirical Bayes" method is used, meaning that the unknown K_i term is integrated out of equation (3.23) to obtain

$$\overline{L}_i = \left[\int_0^{\infty} e^{-K_i\lambda_d(i)t_o(i)} K_i^{\,n_o(i)} \frac{e^{-\rho K_i}}{\Gamma(\delta)} (\rho K_i)^{\delta-1} \rho \, dK_i\right] \left[e^{-\lambda_d(i)t_d(i)} (\lambda_d(i))^{n_d(i)+n_o(i)}\right]. \tag{3.25}$$

The integral in the above expression can be reduced to

$$\frac{\rho^{\delta}}{\Gamma(\delta)} \int_0^{\infty} K_i^{(n_o(i)+\delta)-1} e^{-(\lambda_d(i)t_o(i)+\rho)K_i} dK_i \, . \tag{3.26}$$

Noting that the cumulative distribution function (CDF) of a gamma distribution with parameters $(n_o(i)+\delta)$ and $(\lambda_d(i)t_o(i)+\rho)$ has a known total area of 1.0, the following identity can be stated:

$$\frac{(\lambda_d(i)t_o(i)+\rho)^{n_o(i)+\delta}}{\Gamma(n_o(i)+\delta)} \int_0^{\infty} K_i^{(n_o(i)+\delta)-1} e^{-(\lambda_d(i)t_o(i)+\rho)K_i} dK_i = 1. \tag{3.27}$$

Rearranging equation (3.26) to take advantage of (3.27), equation (3.26) can be reduced to

$$\frac{\Gamma(n_o(i)+\delta)\rho^{\delta}}{\Gamma(\delta)(\lambda_d(i)t_o(i)+\rho)^{n_o(i)+\delta}} \, . \tag{3.28}$$

Therefore, equation (3.25) can be re-written as

$$\bar{L}_i(\lambda_d(i);\rho;\delta;i-\text{data}) = \left[e^{-\lambda_d(i)t_d(i)}(\lambda_d(i))^{n_o(i)+n_d(i)}\right]\left[\frac{\Gamma(n_o(i)+\delta)\rho^{\delta}}{\Gamma(\delta)(\lambda_d(i)t_o(i)+\rho)^{n_o(i)+\delta}}\right]. \tag{3.29}$$

It is generally easier to work with the natural log of the likelihood function

$$\begin{aligned}\bar{\ell}_i = \ln(\bar{L}_i) = {} & -\lambda_d(i)t_d(i) + (n_d(i)+n_o(i))\ln(\lambda_d(i)) + \ln(\Gamma(n_o(i)+\delta)) \\ & + \delta\ln(\rho) - \ln(\Gamma(\delta)) - (n_o(i)+\delta)\ln(\lambda_d(i)t_o(i)+\rho)\end{aligned}. \tag{3.30}$$

The partial derivative of equation (3.30) with respect to $\lambda_d(i)$ is

$$\frac{\partial\bar{\ell}_i}{\partial\lambda_d(i)} = \frac{n_d(i)-\lambda_d(i)t_d(i)}{\lambda_d(i)} + \rho\frac{n_o(i)-\frac{\delta}{\rho}\lambda_d(i)t_o(i)}{\lambda_d(i)(\lambda_d(i)+\rho)}. \tag{3.31}$$

Setting equation (3.31) equal to zero and solving for $\hat{\lambda}_d(i)$ gives the maximum likelihood estimator for $\lambda_d(i)$. Unfortunately, however, the parameters δ and ρ are unknown.

One way to get at the parameters δ and ρ is to obtain an initial estimate for $\lambda_d(i)$ using the formula

$$\hat{\lambda}_d(i) = \frac{n_d(i)}{t_d(i)}, \tag{3.32}$$

and then derive maximum likelihood estimators for δ and ρ. One can then use these initial estimates for δ and ρ in equation (3.31) to obtain an adjusted value for $\hat{\lambda}_d(i)$ for each system. It may be possible to continue these iterations until $\hat{\delta}$, $\hat{\rho}$, and $\hat{\lambda}_d(i)$ all stabilize.

Then, using $\hat{\delta}$ and $\hat{\rho}$, as the parameters for the gamma prior distribution, it is possible to estimate the mean K_i factor and its variance using the relationships

$$E[K_i] \approx \frac{\hat{\delta}}{\hat{\rho}} \tag{3.33}$$

and

$$Var[K_i] \approx \frac{\hat{\delta}}{(\hat{\rho})^2}. \tag{3.34}$$

Assuming data exists on more than one "similar" system, say I of them, then the log-likelihood function becomes, up to irrelevant constants

$$\bar{\ell} = \sum_{i=1}^{I} \bar{\ell}_i = \sum_{i=1}^{I} \left[\delta \ln(\rho) + \ln\left[\frac{\Gamma(n_o(i)+\delta)}{\Gamma(\delta)}\right] - (n_o(i)+\delta)\ln(\lambda_d(i)t_o(i)+\rho) \right]. \tag{3.35}$$

The ratio of gamma functions in equation (3.35) can be simplified by expanding the numerator and then canceling the denominator

$$\begin{aligned} \frac{\Gamma(n_o(i)+\delta)}{\Gamma(\delta)} &= \frac{(n_o(i)+\delta-1)(n_o(i)+\delta-2)\cdots(n_o(i)+\delta-n_o(i))\Gamma(\delta)}{\Gamma(\delta)} \\ &= \begin{cases} 1, & n_o(i)=0 \\ (n_o(i)+\delta-1)(n_o(i)+\delta-2)\cdots(n_o(i)+\delta-n_o(i)), & n_o(i)\ge 1. \end{cases} \end{aligned} \tag{3.36}$$

Then the natural log of equation (3.36) becomes

$$f^*(\delta) = \ln\left(\frac{\Gamma(n_o(i)+\delta)}{\Gamma(\delta)}\right) = \begin{cases} 0, & n_o(i)=0 \\ \sum_{j=0}^{n_o(i)-1} \ln(\delta+j), & n_o(i)\ge 1. \end{cases} \tag{3.37}$$

Rewriting equation (3.35) gives

$$\bar{\ell}=\sum_{i=1}^{I}\left[\delta\ln(\rho)+f^{*}(\delta)-\left(n_{o}(i)+\delta\right)\ln\left(\lambda_{d}(i)t_{o}(i)+\rho\right)\right] \tag{3.38}$$

which can be further simplified to

$$\bar{\ell}=\sum_{i=1}^{I}\left[\delta\ln(\rho)-\left(n_{o}(i)+\delta\right)\ln\left(\lambda_{d}(i)t_{o}(i)+\rho\right)\right]+\sum_{i=1}^{I}f^{*}(\delta). \tag{3.39}$$

It is now a simple matter to take the partial derivatives of equation (3.39) with respect to δ and ρ which, when set equal to zero, will enable one to solve for maximum likelihood estimators for the two parameters. Specifically

$$\frac{\partial\bar{\ell}}{\partial\delta}=I\ln(\rho)-\sum_{i=1}^{I}\ln\left(\lambda_{d}(i)t_{o}(i)+\rho\right)+\sum_{i=1}^{I}\frac{\partial}{\partial\delta}f^{*}(\delta) \tag{3.40}$$

where

$$\frac{\partial}{\partial\delta}f^{*}(\delta)=\begin{cases}0, & n_{o}(i)=0\\ \displaystyle\sum_{j=0}^{n_{o}(i)-1}\frac{1}{\delta+j}, & n_{o}(i)\geq 1\end{cases} \tag{3.41}$$

and

$$\frac{\partial\bar{\ell}}{\partial\rho}=\sum_{i=1}^{I}\frac{\frac{\delta}{\rho}\lambda_{d}(i)t_{o}(i)-n_{o}(i)}{\left(\lambda_{d}(i)t_{o}(i)+\rho\right)}. \tag{3.42}$$

Finally, the preceding two equations, (3.40) and (3.42) along with (3.31) give I + 2 equations with I + 2 unknowns which can be solved to obtain $\hat{\lambda}_{d}(i)$, $\hat{\delta}$ and $\hat{\rho}$.

So far, the author has not been able to solve the system of I+2 equations and I+2 unknowns. Solution with I = 10 similar systems was attempted using the non-linear solver included with Microsoft Excel 4.0 with solver options set as follows: Newton search method, using forward derivatives, with tangential estimates. Moreover, attempts to use the same methodology with only two "similar" systems; i.e., four equations and four unknowns, have also failed. The author believes that equations (3.31), (3.40) and (3.42), or at least some combination of the equations, are ill-conditioned in that they fail to converge at a common point in (I+2) space. An attempt was made to re-parameterize the equations, substituting the variables $\lambda_d(i)$, ρ and δ with exponentials, however this attempt also failed to converge.

While it appears the hierarchical Bayes methodology has promise, as well as a certain intuitive appeal, further research is called for in order to better condition the set of equations needed to use the model. Other issues to consider are whether or not the gamma prior is appropriate for this model and if so, how the set of equations needed to use the model can be simplified or re-parameterized so that reliable model parameter estimates can be obtained.

IV. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

A. SUMMARY

The purpose of this thesis is to (1) investigate the behavior of a statistical procedure used to determine the length of time needed to test a continuously operating system when the system under test is assumed to have an exponential time to failure distribution, but when in fact the system does not have an exponential time to failure distribution and rather may have a time to failure distribution that more closely resembles a Weibull distribution; (2) to examine the behavior of the Gaver-Jacobs estimator for the relationship between DT and OT failure rates using a Monte Carlo simulation; and (3) to introduce a hierarchical Bayes method for estimation of the relationship between the DT and OT failure rates. In the first part of the thesis, the parameters of the procedure are chosen based on the assumption of an exponential time to failure distribution. The behavior of the procedure is then investigated under the assumption that the time between failures are independent, having a Weibull distribution with the same mean time to failure. The intention is to test whether or not use of the Weibull distribution will result in fewer test hours to verify that a system under test has a mean time to failure which meets or exceeds a minimum acceptable value at a given level of confidence, and also to test whether or not the OTAs are misstating the confidence they report in their operational testing results when the exponential distribution is used to model a system whose true failure rate function is that of a Weibull distribution. In the second part of the thesis, the Gaver-Jacobs estimator is examined in terms of the consistency and variability of the OT failure rate derived using that estimator compared to other methods of establishing the OT failure rate. Finally, in the third part of the thesis, a hierarchical Bayes model is proposed for the ultimate purpose of being able to describe the mean and variance associated with a random relationship between DT and OT failure rates.

B. CONCLUSIONS

1. On comparison of the exponential and Weibull failure models.

The first analysis suggests that operational testing agencies can, under certain circumstances, reduce the amount of time needed to demonstrate that the mean time to failure (MTTF) of a continuously operating system is at least as large as the minimum acceptable value (MAV) if the system's true underlying failure distribution is Weibull with shape parameter $\beta > 1$. However, the shape parameter is seldom, if ever, known or even estimable without expending valuable test resources, so it is unlikely that an OTA could actually save any time or resources with this methodology. Moreover, from the perspective of confidence levels, the analysis suggests that confidence in one's result is not sacrificed when the exponential distribution is used to model a system which has a non-constant failure rate, since it appears that use of the exponential distribution on a truly Weibull system seems to cause the OTA to overestimate the number of hours needed to test the MAV for the MTTF to a given confidence level, and also to underestimate the confidence it reports in its results under the incorrect assumption of an exponential distribution. However, for reasons mentioned previously, these results may not be true in general. Therefore, it is concluded that there is currently little to be gained by attempting to model the time between system failures with the Weibull distribution as previously suggested, particularly if very few observations of time to failure are available.

2. On using the Gaver-Jacobs method to combine DT and OT failure data.

The second analysis indicates that the Gaver-Jacobs method for estimating a system's OT failure rate given that system's DT failure rate as well as failure data on previous similar systems produces a reasonable estimate of the actual OT failure rate (when the OT failure rate behaves as expected). In other words, if the relationship between the DT and OT failure rates for a certain system is such that it can in fact be estimated from previous similar systems, then the Gaver-Jacobs method gives a good estimate of the OT failure rate given that system's DT failure rate. In fact, the estimate provided by the Gaver-Jacobs method appears to be nearly as good as direct observation of the OT failure rate obtained through actual testing. Thus, if it can be shown that inferences about the failure rate of a system can in fact be made on the basis of previous similar systems, then the Gaver-Jacobs method of predicting a new system's OT failure rate from previous DT failure data has some potential utility. In particular, this OT failure rate prediction could be used in ways which would aid decisions as to when to begin OT as well as to augment follow-on OT in an effort to reduce needed testing resources.

3. On the hierarchical Bayes Model for estimating K.

The third analysis is incomplete. It was hoped that the hierarchical Bayes method given would provide a means by which to estimate K, the relationship between the DT and OT failure rates, when K is assumed to be a random variable with some prior distribution. Regretfully, the equations derived for the maximum likelihood estimators for the parameters of the gamma prior have proved to be difficult to solve satisfactorily. The author was unable to find a meaningful solution to the set of I+2 equations and I+2 unknowns, where I represents the number of "similar" systems for which previous DT and OT data exist. If such a solution were available it could be used to predict future failure rates by means of a Bayesian calculation. Investigation of such a procedure is left for a later time.

C. RECOMMENDATIONS FOR FURTHER STUDY

Concerning the notion of modeling the failure distribution of a continuous system using the Weibull distribution, OTAs could clearly benefit from being able to use this approach when appropriate. If the time between failures for a system has a Weibull distribution, then proper choice of that system's Weibull shape parameter would probably enable the OTA to reduce the required test resources while maintaining a specified level of confidence in the result of their test. Unfortunately, there is very little data in existence which can assist analysts in estimating the shape parameter of a system unless it is assumed that shape parameter information can be "borrowed" from similar systems. Moreover, it is seldom worth the effort needed to submit the system to special testing just to determine it's shape parameter. Finally, there is some doubt that the necessary relationship between expected time to failure and the numbers of failures observed in a fixed time exists between the exponential and Weibull distributions. Therefore, one area for further research would be in devising methods by which analysts could estimate the shape parameter of a system prior to testing; see the "borrowing information" idea suggested above. Another area of further research, which would be useful in the area previously mentioned, would be in creating a database of system types and the associated Weibull shape parameters of their inter-failure distribution. A third area of investigation would be in identifying those combinations of exponential and Weibull parameters in which a suitable relationship exists between the expected time to failure and the numbers of failures observed in a fixed time.

Concerning the combining of DT and OT failure data, the ability to make inferences about the OT failure rate of a system prior to testing would certainly benefit the OTAs by enabling them to defer testing until they are reasonably sure the system will pass the test. Moreover, the ability to augment future OT failure data with pseudo-data obtained from prior DT failure data would enable the OTAs to reduce the amount of testing resources needed to verify reliability thresholds. To date, the notion of inferring the relationship between DT failure data and OT failure data based on previous similar systems is untested.

Therefore, further areas of research in this field include the investigation of actual data on several similar systems to determine the feasibility of detecting any relationship that might exist between them and also to investigate the nature of the relationship which is found.

Finally, concerning the estimation of the K factor using a hierarchical Bayes methodology, further research is needed in order to develop a more stable set of equations for the parameters of the gamma prior. Moreover, one could consider whether or not a gamma prior is appropriate for this model, or whether some other prior distribution might be more appropriate.

APPENDIX A. ALTERNATIVE METHOD OF DERIVING TIME TO TEST

There is an alternative way to determine, for a given level of confidence, the length of time, t, needed to test a continuous-type system in order to demonstrate that the true MTTF of the system is at least as large as the MAV.[10]

Let the MAV be denoted by θ, the 100α percent lower confidence limit for γ, the true MTTF. The usual test plan is to test for a length of time, t, and observe F, the random number of failures to occur in (0, t]. The planned test time, t, is chosen so that if one failure is observed, the computed lower confidence limit, γ_L, will be equal to the MAV, θ. The lower confidence limit, γ_L, is the solution for γ in the equation

$$1-\alpha = P(T > t) = P(N(t) \le F) = \sum_{i=0}^{F} \frac{\left(\frac{t}{\gamma}\right) e^{-\left(\frac{t}{\gamma}\right)}}{i!} \tag{A.1}$$

where T is the time to the (F+1)st failure.

The following relationship can be used to solve for γ in equation (A.1)

$$\sum_{j=0}^{a-1} \frac{e^{-\lambda}\lambda^j}{j!} = \int_{2\lambda}^{\infty} \frac{1}{2^a \Gamma(a)} x^{a-1} e^{-\frac{x}{2}} dx = P\left(\chi^2_{2a} \ge 2\lambda\right) \tag{A.2}$$

where χ^2_{2a} denotes a chi-square random variable with 2a degrees of freedom. Equation (A.1) then becomes

---

[10] The author is deeply indebted to W. Max Woods, Professor of Operations Research, Naval Postgraduate School, for his guidance in the derivation of this alternative method.

$$1-\alpha = P\left(\chi^2_{2(F+1)} \geq \frac{2t}{\gamma}\right) \tag{A.3}$$

which implies that

$$\frac{2t}{\gamma} = \chi^2_{\alpha,2(F+1)} \quad \Rightarrow \quad \gamma = \frac{2t}{\chi^2_{\alpha,2(F+1)}} \tag{A.4}$$

where $\chi^2_{\alpha,2(F+1)}$ is the αth percentile of the chi-square distribution with 2(F+1) degrees of freedom. Since the solution for γ is γ_L, and the MAV is θ,

$$\theta = \gamma_L = \frac{2t}{\chi^2_{\alpha,2(F+1)}}. \tag{A.5}$$

Therefore

$$t = \theta \frac{\chi^2_{\alpha,2(F+1)}}{2}. \tag{A.6}$$

If $F = 1$ and $\alpha = 0.8$, then $\chi^2_{0.8,4} = 5.999$ and thus, from equation (A.6), $t = 3\theta$, as shown earlier in section II.A. Note that equation (A.6) holds for all positive values of θ and F, making this a very simple way to solve for t given any values of θ and F.

APPENDIX B. CODE FOR RANDOM POISSON ARRIVAL GENERATOR

In this appendix, the Poisson failure generator discussed in Chapter III is given. The program was used to simulate the number of failures during a given phase of testing for a given exponential failure rate and a given test duration. The program was written in Pascal and executed using *Turbo Pascal 6.0* (Borland, 1990), a commercial Pascal package.

```
Program Generate_Failures;

uses TOOLBOX, CRT;

{ Author:  Timothy P. Anderson, Naval Postgraduate School
  Date:    27 March 1994
  Purpose: To generate numbers of failures that occur during a pre specified
        time, TIME, when the underlying failure distribution is
        exponential with rate LAMBDA.  This program is written in partial
        support of a thesis by the author.
}

var
  Lambda, Time, X, CumTime, Rand : real;
  I, N, Count : integer;
  View: boolean;
  OutFile : string;
  Ch : char;

begin
  CRT.Clrscr;
  writeln ('Welcome to Program Generate_Failures... Follow instructions...');
  writeln; writeln;
  write ('Specify the random seed:> '); readln (randseed); writeln;
  write ('Specify the exponential failure rate, Lambda:> ');
  Lambda:= TOOLBOX.Get_Real; writeln;
  write ('Specify the testing time allowed:> ');
  Time:= TOOLBOX.Get_Real; writeln;
  write ('Specify number of iterations to run:> ');
  Count:= TOOLBOX.Get_Integer; writeln;
  writeln ('You may view output on screen or send it to a text file.');
```

```
  write ('Do you want to view it on the screen? (Y/N):> ');
  View := TOOLBOX.Get_Answer; writeln; writeln; writeln;
  if View = False then begin
    write ('Specify the output file (including path):> ');
    read (OutFile); readln;
    SYSTEM.Assign (Output, OutFile);
    rewrite (Output);
  end; {if}

  for I := 1 to Count do begin
    X := 0;
    N := 0;
    CumTime := 0;
    repeat
      Rand := Random;
      X := -(1/Lambda)*ln(1 - Rand);
      if (X + CumTime) > Time then begin
        writeln (N);
        CumTime := CumTime + X;
      end
      else begin
        N := N + 1;
        CumTime := CumTime + X;
      end; {if}
    until CumTime > Time;
  end; {for}

  SYSTEM.Close (Output);
  SYSTEM.Assign (Output, 'con');
  rewrite (Output);

  if View = False then begin
    writeln ('Output sent to ', OutFile,'... Press any key to exit.');
    Ch := Readkey;
  end
  else begin
    writeln ('Press any key to exit.');
    Ch := Readkey;
  end;

end.
```

LIST OF REFERENCES

Barlow, R. E. and Proschan, F., *Statistical Theory of Reliability and Life Testing Probability Models*, Holt, Rinehart and Winston, 1975.

Borland International, *Turbo Pascal 6.0*, 1990.

COTF, COMOPTEVFORINST 3960.7B, *Analyst's Notebook*, June 16, 1992.

Cox, D. R., and Hinkley, D. V., *Theoretical Statistics*, Chapman and Hall, 1974.

Gaver, D. P. and Jacobs, P. A., "On Combining Developmental Testing and Operational Testing Data," Naval Postgraduate School, March 23, 1994.

Keller, C., "Reporting Confidence Levels in Final Test Reports," Center for Naval Analysis, April 14, 1993.

Madson, R. O., "Potential NPS Thesis Topics," COMOPTEVFOR 332H, 1993.

NAVORDSYSCOM, NAVORD OD 44622, *Reliability Guides*, v. 2, "Reliability Test and Evaluation," 1971.

Ross, S. M., *Stochastic Processes*, Wiley, 1983.

INITIAL DISTRIBUTION LIST

		No. of Copies
1.	Defense Technical Information Center Cameron Station Alexadria, VA 22304-6145	2
2.	Library, Code 052 Naval Postgraduate School Monterey, CA 93943-5002	2
3.	Professor Donald P. Gaver Department of Operations Research (OR/Ga) Naval Postgraduate School Monterey, CA 93943-5002	1
4.	Professor Patricia A. Jacobs Department of Operations Research (OR/Ja) Naval Postgraduate School Monterey, CA 93943-5002	1
5.	Professor W. Max Woods Department of Operations Research (OR/Wo) Naval Postgraduate School Monterey, CA 93943-5002	1
6.	Commander, Operational Test and Evaluation Force Technical Director, Code 00T 7970 Diven Street Norfolk, VA 23505-1498	1
7.	Commander, Operational Test and Evaluation Force Head, Operational Analysis, Code 35 7970 Diven Street Norfolk, VA 23505-1498	1
8.	LT Timothy P. Anderson 4813 East Dolphin Avenue Mesa, AZ 85206	1